METRON

RIVISTA INTERNAZIONALE DI STATISTICA — REVUE INTERNATIONALE DE STATISTIQUE
INTERNATIONAL REVIEW OF STATISTICS — INTERNATIONALE STATISTISCHE ZEITSCHRIFT
REVISTA INTERNACIONAL DE ESTADISTICA

DIRETTORE PROPRIETARIO — DIRECTEUR ET PROPRIÉTAIRE
EDITOR AND PROPRIETOR — HERAUSGEBER UND EIGENTHÜMER
DIRECTOR Y PROPIEDARIO

Prof. Dott. Corrado Gini, *Direttore dell'Istituto di Statistica della Università di Roma*

AMMINISTRATORE — ADMINISTRATEUR — MANAGER — VERWALTER — ADMINISTRADOR

Dott. Franco Mariani, *Istituto di Statistica della Università di Roma*

SEGRETARI DI REDAZIONE — SECRÉTAIRES DE RÉDACTION
EDITORIAL SECRETARIES — REDACTIONSSECRETAERE — SECRETARIOS DE LA REDACCIÓN

Proff. V. Castellano, N. Federici, E. Pizzetti, G. Pompilj

Vol. XVII - N. 1-2 **31-XII-1953**

SOMMARIO-SOMMAIRE-CONTENTS-INHALT-SUMARIO

Corrado Gini. *La tyrannie des chiffres* . . . . . pag. 3
Corrado Gini. *Di alcuni simboli che può essere opportuno impiegare nella trattazione dei fenomeni statistici.* . . . . . » 25
S. R. Das. *A mathematical analysis of the phenomena of human twins and higher plural births. Part I : Twins* . . . . . » 65
G. Herdan. *Language in the light of the theory of information* . . . . . » 89
Corrado Gini. *Le basi statistiche delle diagnosi e delle prognosi* . . . . . » 127
Carlo Benedetti. *Di un confronto fra poteri d'acquisto di salari e considerazioni metodologiche relative agli indici circolari e non circolari* . . » 151

ROMA

AMMINISTRAZIONE DEL « METRON »

UNIVERSITÀ — ISTITUTO DI STATISTICA

ARTICOLI GIUNTI ALLA RIVISTA CHE VERRANNO PUBBLICATI NEI PROSSIMI NUMERI (*secondo l'ordine di arrivo*).

ARTICLES RECEIVED BY THE REVIEW WHICH WILL BE PUBLISHED IN FUTURE ISSUES (*according to date of receipt*).

ARTICULOS LLEGADOS A LA REVISTA QUE SE PUBLICARÁN EN LOS PROXIMOS NUMEROS (*según el orden de su llegada*).

ARTICLES REÇUS PAR LA REVUE ET À PARAÎTRE PROCHAINEMENT (*d'après la date de reception*).

ARTIKEL DIE AN DIE ZEITSCHRIFT ANGELANGT SIND UND WELCHE IN DEN FOLGENDEN NUMMERN ERSCHEINEN WERDEN (*nach der Reihenfolge des Eingangs*).

C. GINI. — *Probabilità che x termini di una serie erratica siano tutti decrescenti o tutti crescenti.*

G. HERDAN. — *Language in the Light of the Statistical Theory of Information. Part. II.*

S. R. DAS. — *A mathematical analysis of human twins.*

Part II : *Triplets and the Application of the analysis in the Interpretation of the Twin and the Triplet Data.*

Part III : A. *Correction of Data for prenatal Mortality.*
B. *Refinement of the Theory for Decline of the Twinning Processes with Increase of Plurality.*

C. BENEDETTI. — *Del massimo valore dell'indice di oscillazione in una successione di termini al variare in tutti i modi possibili l'ordine di questi.*

G. TINTNER. — *The Distribution of the Variance of Variable Differences in the Circular Case.*

C. BENEDETTI. — *Sul massimo e minimo dell'indice semplice di dissomiglianza.*

Gli Autori degli articoli inviati per la pubblicazione nella Rivista, rinunciano in favore della medesima alla proprietà letteraria degli articoli stessi, qualora vengano pubblicati.

Les Auteurs des articles envoyés à la Revue, pour y être publiés, renoncent, en faveur de celle-ci, à la proprieté littéraire de leurs articles, s'ils sont acceptés.

The Authors of papers sent for publication in the Review are supposed to give up their copyright in favour of the Review if the papers are published.

Die Verfasser der zur Veröffentlichung in der Zeitschrift zugesandten Aufsätze, werden, falls selbige veröffentlicht werden, auf ihre Verfasserrechte zu Gusten der Zeitschrift verzichten mussen.

Los Autores de los artículos enviados para su publicación en « Metron » renuncian su propriedad a favor de la Revista cuando los artículos sean publicados.

CORRADO GINI

La tyrannie des chiffres

« Dieu géométrise » — écrivait Platon, voulant entendre par là que l'univers tout entier est un tissus de rapports quantitatifs ; « l'homme géométrise », — pourrait-on dire aujourd'hui, car il a tissé de chiffres la trame de sa vie. Chiffres qui augmenteront encore à l'avenir, si nous en croyons l'exemple des Etats-Unis d'Amérique, qui, par bien d'aspects, devancent l'évolution de l'humanité.

En Amérique — aux Etats-Unis, voulons-nous dire — tout est gradué, énuméré, évalué, mesuré.

On établit des gradations pour les champions sportifs, les chefs d'orchestre, les stars du cinéma ; on établit des gradations pour les beautés féminines par des concours locaux, professionnels, régionaux, nationaux, mondiaux ; une gradation pour les criminels (et les journaux nous informent que le sceptre de la criminalité a passé des mains de Al Capone à celles de Lucky Luciano, puis de Frank Costello, enfin de Thomas Lucchese) ; une gradation pour les Universités, arrêtée à la suite des plébiscites entre les licenciés (Harvard se trouve au premier rang, suivie de près par Chicago) ; une gradation pour les livres d'après leur diffusion ; une gradation pour les ponts, les gratte-ciels, les paquebots, les gares, les ports, les parcs, les sociétés, les villes, les fleuves, les grands revenus et les grosses fortunes.

Le classement ne suffit d'ailleurs pas ; on évalue tout, la plupart du temps en dollars, comme pour les revenus et les fortunes d'importance, les mouvements des ports et des grandes compagnies ; parfois, si possible, aussi suivant le temps exigé par l'édification de l'ouvrage, ou la hauteur ou la longueur de celui-ci, comme pour les gratte-ciels et les ponts et, en général, pour les constructions ; ou bien encore suivant le nombre d'habitants, comme pour les villes, ou la superficie, comme pour les bassins des fleuves ou pour les parcs publics.

Que ces habitudes soient bonnes ou mauvaises, elles gagnent du terrain et quelques-unes sont déjà répandues dans le monde entier, comme c'est le cas pour les concours des reines de beauté ou pour le classement des champions sportifs.

Mais, dans d'autres domaines, l'usage des chiffres s'est généralisé depuis longtemps désormais, et est devenu presque un signe de civilisation.

Le premier de ces domaines a été peut-être celui de la surface terrestre, lorsque l'homme, renonçant à la vie nomade, a choisi un lieu d'habitation stable et s'est mis à cultiver la terre. La profession d'expert dans la mesure des champs a été l'une des plus importantes professions à l'aube des civilisations, et le dieu Terme était si respecté à Rome que son effigie se trouvait dans le Temple de Jupiter Capitolin. Plus tard, les Itinéraires représentèrent l'une des principales expressions de l'organisation de l'Etat. Graduellement, la mesure des distances se fit plus précise et devint plus répandue. Il ne s'agissait plus de chiffres ensevelis dans des archives, mais de données portées à la connaissance du public. D'abord, les distances furent marquées sur quelques bornes placées de temps en temps le long des routes ; puis on prit l'habitude de mettre ces bornes tous les kilomètres et, actuellement, aux bornes kilométriques s'alternent, sur les routes les plus importantes, des signaux tous les cent et même tous les dix mètres.

La mensuration du temps remonte peut-être à des époques plus lointaines encore, et occupe une place tout aussi considérable dans la vie des peuples. Bornée d'abord aux phénomènes célestes et confiée aux astronomes, pour qu'ils déterminent les dates des moments les plus saillants dans la succession des saisons et la rencontre des astres, elle s'étendit graduellement à la vie quotidienne, en réglant le rythme avec une assiduité et une sévérité croissantes. Les hommes se contentèrent, d'abord, de partager la journée suivant les phases de la lumière solaire, les mesurant au moyen des cadrans solaires ; ceux-ci étaient complétés, mais seulement dans certains cas, pour les petites nécessités de l'existence, par les sabliers.

Puis, le son des cloches des beffrois ou la voix du *muezzin* du haut des minarets rappelèrent aux fidèles la naissance et la clôture de leur journée de labeur, réglée encore une fois selon les déplacements de Phébus ; pour les catholiques, la cloche mar-

qua aussi l'heure du repas de midi et parfois même — comme dans les Retraites des Jésuites du Paraguay — elle régla, au cœur de la nuit, l'accomplissement de devoirs plus intimes.

La construction de l'horloge marqua un grand progrès, car elle permit de calculer le temps indépendamment des changements atmosphériques, et de façon ininterrompue ; symbole de la réglementation sociale, l'horloge domina, du haut de l'Hôtel-de-Ville, ou de l'Eglise, du beffroi ou du clocher, et l'ingéniosité humaine s'efforça de fabriquer les mécanismes les plus compliqués, les plus délicats, pour marquer, non seulement les heures, les minutes et les secondes, mais aussi les cycles des jours de la semaine et des mois, les signes du zodiaque et d'autres phénomènes célestes ; à cela les artisans trouvèrent moyen de relier des signaux acoustiques ou automatiques. Les battements, ou coups, limités au debut à l'heure, marquèrent bientôt aussi la demi-heure, puis le quart, rythmant ainsi, à des intervalles de plus en plus courts, la vie humaine. Jugeant insuffisantes les marques du temps qui lui venaient des principaux édifices appartenant à la collectivité, l'homme voulut enfin des pendules dans sa demeure, des montres sur sa personne. Les montres, reléguées d'abord dans le gousset ou dans la poche, furent ensuite adaptées à des bracelets, pour être plus facilement consultées. Les horloges passèrent de l'église et de l'Hôtel-de-Ville à la gare, à l'auto, au tramway ; enfin le cadran aussi a été éliminé, et on a adopté l'horloge parlante du téléphone. Le grand nombre d'horloges, dans une ville, est un indice de la civilisation de ses habitants, et la possession d'une montre marque une limite entre les classes sociales supérieures et inférieures ; pour l'individu des classes supérieures, la possession de la montre marque le passage de l'adolescence à la maturité.

Les primitifs se moquent de la prétention des civilisés de connaître la date de naissance de chaque individu. De nos jours, en Europe, les enfants des premières classes et des colonies de vacance déclinent avec désinvolture, avec le nom de leurs père et mère, leur âge. Les recensements, avec la tendance plus ou moins marquée à arrondir le chiffre de l'âge déclaré, sont un indice du progrès dans la mensuration de l'existence et de ses variations dans l'espace, par rapport au développement de la civilisation.

Bien des gens connaissent, maintenant, non seulement leur âge, mais aussi leur taille et beaucoup d'eux savent aussi leur poids ;

d'aucuns sont au courant de la fréquence de leur pouls et de leur température. Le jour viendra, probablement, où chaque individu saura décliner, non seulement sa taille, son poids, ses pulsations, sa température, mais aussi le rythme de sa respiration, sa pression sanguine, son taux hémoglobinométrique et d'autres traits physiologiques.

De même il est probable qu'on finira par connaître couramment, comme cela advient déjà dans les campagnes, la distance entre sa propre habitation ou son propre village, et les localités ou les villes plus ou moins voisines ou éloignées ; ainsi que la population des centres ruraux ou urbains les plus proches, et l'altitude de son propre lieu de résidence et de la région environnante.

La vulgarisation et la précision de ces notions quantitatives sont évidemment d'une grande importance pour le statisticien car la statistique est, sinon surtout, du moins certainement avant tout, ainsi qu'on l'a définie, la « *science de l'enumération* » et, naturellement, sa tâche est d'autant plus facile que les notions quantitatives deviennent des connaissances plus répandues, ainsi que peuvent bien s'en rendre compte les personnes qui ont été chargées de relever des statistiques, non seulement parmi des populations civilisées, mais aussi parmi des populations primitives.

Mais les conséquences sociologiques n'en sont pas moins importantes.

L'habitude d'établir une gradation des objets ou des personnes provoque ou accroît l'émulation entre les intéressés. Le désir de se placer au premier rang fait trouver aux hommes de curieux stratagèmes. C'est ainsi que les constructeurs du Crysler Building, le gratte-ciel qui, à New-York, détenait, avant la construction de l'Empire State Building, le record de hauteur, pour obtenir ce dernier contre le Manhattan Building, qu'on bâtissait simultanément, feignirent d'avoir achevé leur ouvrage, laissant ainsi leurs concurrents terminer complètement le Manhattan ; dès qu'un toit fut placé sur celui-ci, une aiguille haute et pointue fut construite sur le Crysler, qui atteignit ainsi une hauteur supérieure. Au lendemain de la première guerre mondiale, la Grande-Bretagne et les Etats-Unis se partagèrent le butin de deux gigantesques transatlantiques allemands, qui furent rebaptisés avec les noms, l'un de « Majestic » et l'autre de « Leviathan ». Le « Majestic » avait un jaugeage de quelque peu su-

périeur, et les Américains ne s'en consolaient point. Pour battre ce record, il couvrirent les ponts du « Leviathan », si bien que, conformément aux conventions réglementant l'évaluation du jaugeage, ce dernier passa à la tête du classement mondial.

D'ailleurs, le record ne suscite pas seulement des aspirations désintéressées ; ceux qui le détiennent, qu'il s'agisse de biens ou de personnes, exigent qu'on les paie à des prix-record (eux aussi). C'est ainsi que Babe Ruth, champion N. 1 de base-ball, demanda et obtint une rémunération supérieure à celle du Président des Etats-Unis ; on sait aussi que les N. 1 des chefs d'orchestre ou des musiciens, des boxeurs ou des toréros, des chanteurs ou des vedettes ou des metteurs en scène, des gangsters comme des reines de beauté, encaissent des sommes fabuleuses.

Mais ce qui est plus important que la position dominante des recordsmen, dans le domaine des relations sociales, ce sont les restrictions de la liberté des individus, découlant des mensurations précises et des réglementations du temps, de l'espace, du poids, etc..

Le citoyen peut employer les tramways, les autobus, les avions, les trains, les téléphones, les télégraphes, mais seulement à des heures bien arrêtées ; il peut entrer librement dans des jardins et des parcs, des églises, des musées, des théâtres, mais, ici aussi, seulement suivant des horaires ou même des jours établis ; bien des jours — parfois ceux où il lui serait plus aisé de les visiter — ils sont fermés ; les bureaux des administrations publiques suivent, eux aussi, un horaire limité, pour les visiteurs ou les solliciteurs. Pour participer à des concours, présenter des demandes, porter plainte, remettre des listes, des réclamations, etc., pour payer les impôts ou toucher de l'argent, il existe des délais précis et indérogeables. Les mesures législatives n'échappent point à cette limite ; lorsque, le terme expire, toutes les démarches faites, tout le travail accompli, quelque peine qu'ils aient coûté, ne servent plus à rien, et doivent être recommencés de fond en comble. Les droits sont sujets à des prescriptions. Pour le service militaire et les autres rappels sous les drapeaux, les citoyens doivent se présenter à des dates fixes, et les retards sont punis par de graves sanctions. Le développement physique de l'individu ne compte pas ; ce qui compte exclusivement, c'est la date de naissance. Dans le service militaire, la possibilité de tenir compte en quelque sorte du développement individuel

est pourtant admise, et l'incorporation peut être ajournée d'un ou deux ans après la présentation ; dans les écoles primaires, à l'encontre, même cette possibilité est exclue, tout au moins dans les pays que je connais. Durant les premières classes, garçons et fillettes ayant une maturité intellectuelle très différente sont ainsi obligés de suivre tous le même enseignement, qui peut ne rien apprendre de nouveau à d'aucuns, alors que d'autres sont tout-à-fait incapables de le suivre. Un règlement analogue vaut aussi bien — et souvent aussi sans la possibilité de renvoi — pour la mise à la retraite. De même, malheur au Don Juan qui jugerait de la maturité d'une jeune fille d'après son développement physique plutôt que d'après sa date de naissance. Les règlementations concernant les poids admis sur les trains et les avions sont minutieuses ; les valises ne doivent pas dépasser un certain poids ; sur certains tramways, on ne les accepte pas du tout. Il est des compétitions sportives dans lesquelles les hommes et les chevaux ne peuvent pas dépasser certaines limites de poids.

Des restrictions rigoureuses sont imposées par la mesure des distances. On ne peut construire ni planter d'arbres qu'à une certaine distance des autres maisons ou propriétés ; on ne peut pêcher qu'à une certaine distance des ports ; on ne peut chasser ou tirer qu'à une certaine distance des agglomérations.

Les modalités numériques constituent ainsi un épais réseau de restrictions tout autour des agissements de l'homme civilisé.

Théoriquement, je suis un citoyen parfaitement libre ; et libre donc, en théorie, de déjeuner, de sortir de chez moi, de faire des achats dans les magasins quand bon me semble ; mais, pratiquement, pour déjeuner, je dois attendre la distribution du lait, qui advient à une heure donnée ; pour sortir de chez moi, je dois attendre que les moyens de transports en commun commencent à fonctionner, à une certaine heure ; pour faire mes achats, je dois tenir compte des horaires des magasins. Je suis libre d'aller à bicyclette, mais en ville je ne peux point y aller dans certaines rues, ou bien je puis y passer seulement à certaines heures ou dans une seule direction ; et, si j'achète une auto, je dois me soumettre à des règlements analogues ; en outre, je ne peux garer ma voiture qu'à certains endroits, ni dépasser une certaine vitesse. Je suis libre de mes actions, mais, en réalité, je dois aller à mon bureau à telle ou telle heure. A une certaine heure, je ne puis plus ouvrir mon poste radio ou chan-

ter, ou faire des bruits qui dérangeraient mes voisins. Je n'ai qu'un court délai pour révéler que j'ai engagé une femme de ménage, ou pour présenter au fisc le détail de mes revenus. J'ai un petit jardin où j'aimerais tant me délasser à l'ombre, mais je ne puis y planter d'arbres, pour obtenir cette ombre, parce que je dois respecter la distance prescrite.

En réalité, ma liberté est fort limitée. Il est vrai, par ailleurs, que je puis choisir entre l'une ou l'autre activité professionnelle, mais au fond, il s'agit de choisir entre l'une ou l'autre forme de semi-esclavage.

Mais cet usage des nombres — pourra-t-on m'objecter — cette tyrannie, ainsi que vous l'appelez, des chiffres, n'est après tout que l'expression des complications de la vie sociale moderne. Plus le mécanisme est compliqué, plus, pour bien fonctionner, il doit être précis. L'axe d'un char-à-banc danse dans la roue, et le char avance tout de même ; mais les petites roues d'un chronomètre doivent constituer un engrenage d'une précision infinitésimale. De même, le filet des limitations quantitatives de notre vie a des mailles beaucoup plus serrées en ville qu'à la campagne, dans les métropoles que dans les colonies. La ponctualité, qui était autrefois, ainsi que disait Louis XIV, « la courtoisie des rois », est devenue, dans la vie moderne, le devoir de tous les citoyens, et est considérée comme un trait caractéristique du progrès de la civilisation. La ponctualité que le fascisme avait apportée aux chemins de fer italiens fut considéré à l'étranger — d'une façon un peu élémentaire, si l'on veut — comme un indice des progrès réalisés par ce régime, dans le domaine de l'ordre social.

Or, tout cela est certainement vrai : toute dictature, toute tyrannie — aussi bien celle des chiffres que les autres — n'est pas le fruit du caprice ou de la folie, mais la conséquence de nécessités du moment ou du milieu social intérieur ou extérieur. Mais tout cela n'empêche pas qu'elle engendre un sacrifice. C'est surtout lors des phases de passage de l'évolution sociale qu'on le sent.

En 1934, j'eus l'occasion de visiter, pour la première fois, l'Irlande, et, après un séjour à Dublin, je voulus me rendre à Belfast. A cet effet, je me renseignai sur l'horaire des trains et leur ponctualité. Mon interlocuteur me répondit que ce train là

était ponctuel ; que c'était, même, le seul train ponctuel de l'Irlande. Et il ajouta, à ce propos, une anecdote : un étranger ayant demandé à un porteur l'heure de départ d'un train, celui-ci répondit : « A 16 h., et, prenez garde, à 16 h. précises... » et il poursuivit, d'un air désolé : « ... c'est un gros inconvénient pour les voyageurs ! » Effectivement, il avait raison : la ponctualité des chemins de fer est un sérieux inconvénient pour les voyageurs habitués à arriver en retard.

L'année précédente, dirigeant, au Mexique, une expédition scientifique, j'avais eu l'occasion d'expérimenter, non seulement le peu de ponctualité qui y régnait, même parmi les personnes les plus évoluées, mais aussi la grande indulgence du public à ce sujet. A tel point que, par exemple, on considérait excessivement sévère le reproche adressé à une personne qui s'était fait attendre pendant plus d'une demi heure, par une caravane d'une vingtaine de personnes, au moment du départ.

Retourné au Mexique en 1947, j'y ai constaté, à ce point de vue et à bien d'autres, des progrès considérables. Il en a été probablement de même aussi en Irlande.

En Italie, pays qui, par sa ponctualité, n'est peut-être pas à la queue, mais certainement pas, non plus, à la tête des nations civilisées, il arrive souvent de voir, dans la rue, des jeunes filles qui courent — les femmes, on le sait, sont plus proches de la nature que les hommes, et se soumettent donc moins facilement aux règlementations sociales — parce qu'elles sont évidemment en retard. D'ailleurs, ce n'est pas seulement en Italie, mais dans tous les pays civilisés, que la vie se déroule sous la pression des horaires, dans un état de hâte anxieuse. Je ne dirai point que c'est là quelque chose de très raisonnable ; c'est plutôt le fruit d'un instinct mûri par la sélection sociale. Vous connaîtrez probablement l'histoire de la conversation entre l'aviateur et le Bédouin : « Vois-tu, disait l'aviateur, j'arrive, en volant, en quatre heures, là où, avec ton chameau, tu arrives avec peine en quinze jours. — Eh bien, répondit le Bédouin, que fais-tu pendant les 14 jours et quelque restants ? — Oh ! je fais un tas de choses, dit l'aviateur — Et toujours à la même vitesse ? demanda encore le Bédouin — Bien sûr, répondit l'aviateur — Mais pour quoi ? s'enquit l'Arabe — Pour faire d'autres choses encore, ensuite, dit l'aviateur — » Et le Bédouin de conclure : « Ne t'inquiète donc pas ! Toi, en haletant, moi en res-

pirant tranquillement, nous arriverons tous deux au même but inévitable » !

Même dans les pays les plus avancés, où la hâte du travail est la plus marquée, on ressent fortement la contrainte de la vie sociale. Nous en avons un témoignage dans l'irrésistible aspiration aux vacances qui s'est emparée de toutes les couches de la société. Non point que l'on veuille des vacances pour ne rien faire ; souvent, même, on emporte avec soi de quoi travailler, sans nécessité aucune. Mais c'est qu'alors, on peut se consacrer à sa tâche sans impositions d'horaires, sans restrictions, sans limites. C'est la raison qui nous fait choisir un coin mort à la campagne ou à la montagne, aussi éloigné que possible de la société civilisée.

C'est pour la même raison que tant de gens se sentent à leur aise aux colonies, et que ceux qui y sont habitués ne savent plus se réadapter aux obligations de la métropole. Et c'est encore ainsi que s'explique l'impossibilité du sauvage à s'adapter à la vie de notre société. Nous avons l'impression qu'il a été entièrement assimilé. Il porte nos mêmes vêtements, il se lave, soigne ses ongles, se peigne, souvent avec une coquetterie presque féminine. Il apprend à lire, à écrire, à calculer ; il apprécie le théâtre ; il adore le cinéma ; il se sert d'une montre, il respecte les convenances. Mais s'il lui arrive de se retrouver en contact avec son monde, il jette le complet-veston et la montre aux orties, et se replonge avec joie dans la liberté sans contrôle de sa vie primitive.

Si la bicyclette d'abord, l'auto ensuite et enfin la TSF ont remporté un succès si éclatant et immédiat, c'est, je crois, parce qu'elles nous rendent indépendants des horaires des transports en commun et des représentations publiques ; mais elles ont dû, elles aussi, ensuite, s'assujettir à des restrictions multiples, pour être insérées dans l'organisation rigoureuse de la société civilisée.

Il y a lieu, par ailleurs, de souligner combien l'homme s'adapte à la tyrannie des chiffres-adaptation plus ou moins lente, mais, à la fin, totale. Dans peu de domaines autant que dans celui-là, le vieil adage qui veut qu'on s'adapte bon gré mal gré à ce qui est imposé par la force, s'avère juste et vrai.

Le manque de ponctualité, que nous avons signalé pour certains pays, se rapporte généralement à certains domaines de

l'activité humaine et disparaît progressivement, pour céder le pas à un respect presque religieux des délais établis, des « termes ».

Le dieu Terme que les Anciens avaient préposé à la garde des espaces, détermine tout aussi sévèrement aujourd'hui la division du temps.

Les membres d'un parlement s'acharnent sur la présentation d'un projet de loi, mais il suffit que la session s'achève sans que le projet ait été présenté pour que cessent, non seulement toute lutte — ce qui est naturel — mais aussi toute récrimination. Tous s'inclinent devant le dieu Terme, et jusqu'à la session suivante, nul ne parle plus de la nouvelle loi.

En Italie, durant la triste période de l'épuration, le Gouvernement avait cru bon de se débarrasser — par une mesure extraordinaire — de bon nombre de hauts fonctionnaires qu'il considérait coupables de fascisme factieux ; mais, par une erreur matérielle, il calcula les mois comme ayant tous 30 jours, et non pas suivant leur longueur réelle, ce qui fit qu'il prit cette décision le 61ème jour de ses pouvoirs extraordinaires, alors que ceux-ci lui étaient accordés pour une durée de 60 jours seulement. La mesure ne fut donc pas appliquée, et cela était justifié du point de vue juridique ; mais, alors que le Gouvernement résista contre d'autres actes d'indulgence ou même d'équité également justifiés, s'efforçant, pour des raisons politiques, de donner une nouvelle interprétation de la loi, cette fois-là, il ne fit pas la moindre tentative en ce sens, s'inclinant aux vouloirs péremptoires du dieu Terme.

Si, auprès d'une population, la mathématique — la science des nombres — se développe singulièrement, par rapport aux autres activités intellectuelles, et particulièrement si cette circonstance accompagne un régime totalitaire, la tyrannie des chiffres peut atteindre des manifestations impressionnantes.

Nous pouvons avoir une idée du niveau atteint par les Mayas dans leurs calculs mathématiques, grâce à la perfection du calendrier qu'ils avaient élaboré. Ce calendrier était différent et beaucoup plus exact que celui auquel les civilisations occidentales modernes sont parvenues après de pénibles progrès, et dont elles ne savent pas s'émanciper encore. Mais le peuple Maya était devenu le véritable esclave de son calendrier. Ce n'est pas moi qui le dis, mais ce sont là les mots mêmes que

l'on peut trouver dans une récente et intéressante publication qui vient d'être traduite, tout récemment, en Italien, sur les civilisations mortes (1).

J'en extrais quelques autres phrases qui donnent une idée éloquente de la situation qui s'y était instaurée :

Parmi les Mayas — et également parmi les Aztèques qui en héritèrent la culture, et probablement parmi les Toltèques, qui la préparèrent — la science du calendrier dirigeait toutes les activités (2).

En Italie, sous le régime fasciste, nous souffrions à contre-cœur les contraintes que Mussolini imposait au programme des travaux publics, les réglant de façon que l'inauguration en eût lieu pour l'anniversaire de la fondation de Rome ou de la Marche des Fascistes sur la capitale italienne. Mais que dire alors de ce qui se passait chez ces peuples ? Toutes leurs pyramides, par exemple (les trois peuples en questions construisirent des pyramides) étaient bâties suivant l'astronomie et assujetties à la tyrannie du calendrier (3).

Et souvent, autour d'une pyramide existant déjà, ils en élevaient une seconde, lorsqu'ils voulaient immortaliser une nouvelle époque de leur calendrier (4).

D'ailleurs, ce ne sont pas seulement les pyramides, mais toutes leurs gigantesques constructions, qui étaient bâties, non pas lorsqu'elles étaient nécessaires, mais uniquement lorsque le calendrier l'ordonnait (5). (Toutefois, il est probable que cette façon d'agir avait si bien pénétré la psychologie du peuple que celui-ci n'éprouvait réellement la nécessité de ces constructions que lorsque le calendrier le voulait). Et cet usage fut respecté, pendant des siècles, avec une régularité des plus strictes, ainsi qu'on peut le voir d'après les dates gravées sur les monuments — les seules interruptions étant dues à une transmigration ou un fléau (6).

---

(1) C.W. Ceram : *Götter, Gräber und Gelehrte, Roman der Archäologie*, Rowohlt, Hamburg-Stuttgart — Edition italienne : *Civiltà Sepolte*, Einaudi, Turin, 1953 — pp. 414, 420. Ces citations de pages, ainsi que les suivantes, se refèrent à l'édition italienne.

(2) *Ibidem*, p. 373.

(3) *Ibidem*, p. 447.

(4) *Ibidem*, pp. 414, 420.

(5) *Ibidem*, p. 420.

(6) *Ibidem*, p. 420.

Bien plus : sur chaque monument, pas un relief, un dessin un animal, n'était sans quelque rapport direct avec une date. Chacune des construtions Mayas était un calendrier pétrifié (1). Les visages horribles, qui représentent des nombres et désignent des jours, des mois, des périodes, se retrouvent sur les façades, sur les colonnes, dans les frises, sur le rampes des escaliers, des temples et des palais (2). Aucun signe, aucune image, aucune sculpture ne reste étrangère au rapport avec le nombre astronomique. Tout exprime des chiffres et du temps : deux croix sur les paupières d'une tête de serpent, les griffes d'un jaguar sur l'oreille du dieu Kuklulkan, la forme d'un portique, les innombrales « yeux à tau », l'immanquable dessin des escaliers (3).

Il y a plus encore : toute la vie de ces peuples se déroulait sous l'épée de Damoclès des grands cycles de leur calendrier ; par exemple, ils croyaient que le monde finissait tous les 52 ans. D'où la puissance des pontifes, qui, seuls, pouvaient éloigner cette terrible calamité : ils le firent par des moyens de plus en plus durs et cruels, qui, au fur et à mesure, dégénérèrent en atroces sacrifices humains (4). C'est ce qui devait pousser les races antagonistes à la rébellion, que Cortès sut si bien exploiter dans sa conquête de la Nouvelle Espagne.

Le niveau atteint par les Mayas, en fait d'astrologie et de mathématique, sur le Nouveau Continent, fait pendant à celui qu'avaient atteint, sur le Vieux Continent, les Egyptiens et les Sumères, d'où proviennent, à travers Babylone et l'Assyrie d'une part, Crète et les races helléniques, de l'autre, les sources de la science et de la civilisation occidentales. En Egypte il y avait une déesse des Nombres, et ceux qui savent comment les dieux égyptiens ne plaisantaient pas, se rendent bien compte de combien ses ordres devaient être péremptoires.

A Babylone la tyrannie des chiffres se manifesta en superstitions caractéristiques dont le legs parvint jusqu'à nous.

C'est à la science de Babylone que s'abreuva Pythagore, que l'histoire rappelle, avec son école, non seulement pour les progrès qu'il fit faire aux mathématiques, mais aussi à cause des vertus magiques attribuées au nombres.

---

(1) C. W. Ceram, *op. cit.*, p. 412.
(2) *Ibidem*, p. 415-416.
(3) *Ibidem*, p. 439.
(4) *Ibidem*, p. 447.

Plus tard, toutes les études médiévales en sont pleines et toute la littérature s'en ressent. D'ingénieuses recherches ont été faites, chez nous, par le Prof. Benini, pour révéler les formules numériques sur lesquelles se serait basée la trame originelle de la « Divine Comédie » de Dante.

De curieuses superstitions numériques survivent, même chez des personnes éminentes, et leur diffusion est grande dans le peuple. D'Annunzio refusait d'écrire le 13 (puisqu'on affirme généralement qu'il porte malheur) et le remplaçait par 12 + 1. Le 13 et de même le 17 (qui aussi est accusé de porter malheur) sont souvent omis dans la numération des chambres des hôtels. Dans les certificats de naissance élaborés dans certaines régions d'Italie, un collègue particulièrement attentif a remarqué, en répondant à un questionnaire sur les déclarations retardées (que j'avais envoyé par l'entremise de l'Institut central de Statistique, pour les travaux de la commission sur la natalité) que le nombre desdites déclarations était sensiblement inférieur — surtout dans le Midi — les 17 de chaque mois et même à 17 heures. Par contre, dans d'autres pays, en particulier en Tchécoslovaquie, on déplace aux premiers jours du mois (et cela non seulement au mois de janvier, pour des raisons d'utilité déjà amplement analysées) la date des naissances survenues pendant les derniers jours du mois précédent.

A part ces manifestations exceptionnelles, les chiffres exercent indéniablement un pouvoir suggestif plus fort que les arguments et les qualifications, mais pas toujours dans le sens le plus approprié. Cela vient de ce que j'appellerai leur *nudité*, ainsi que de leur *comparabilité*.

Nous pouvons distinguer un argument bon d'un argument mauvais, un argument fort d'un argument faible, mais un chiffre faux ou approximatif est identique à un chiffre exact.

Un argument peut être ressemblé à un homme correctement habillé, dont l'uniforme ou la tenue revèle au premier abord, par ses emblèmes, sa propriété ou son élégance, le rang ou la qualité de la personne ; un chiffre ressemble à un homme tout nu, qui n'offre aucune de ces caractéristiques différentielles. C'est pour cela que je parle de *nudité* des chiffres, et c'est pourquoi j'ai toujours insisté, et j'insiste, sur la nécessité que tout chiffre soit toujours accompagné des pièces justificatives, même lorsqu'il est mis en circulation par les statisticiens les plus autorisés

et surtout lorsqu'il est le résultat d'évaluations ou de calculs à caractère plus ou moins subjectifs. C'est exactement comme si l'homme nu dont nous parlions était affiché d'un placard qui permettrait de juger de sa personnalité.

Une autre cause de la suggestion, non sans inconvénients, exercée par les chiffres, est leur comparabilité .

Vous ne pouvez pas dire que la bonté, l'héroïsme, la discipline d'une personne sont de 10 ou de 20 pour cent supérieurs ou inférieurs à ceux d'une autre ; mais vous pouvez le dire de leur taille, de leur poids, de leur revenu ou de leur fortune. Si nous voulons faire des comparaisons, nous devons forcément donner la préférence aux qualités chiffrables.

Il est vrai que de la sorte on peut négliger des côtés importants et assez décisifs pour une appréciation relative des objets en question, et cela est souvent porté comme un argument contre la statistique. Lord Balfour a déjà affirmé que l'amour, la bonté et le bonheur, les trois plus grandes choses au monde — disait-il — ne sont pas susceptibles de mesure. Pour l'amour et la bonté cela n'est pas exact, ainsi que je l'ai fait remarquer ; mais fût-il même exact, est-ce que nous devrions par conséquent renoncer à mesurer et à comparer les qualités mesurables et comparables ? Evidemment non. Ce que nous devrions faire pourtant c'est de nous rendre compte — et les statisticiens les plus avisés s'en rendent compte (je citerai pour tous M. Idenburg dans son discours *Statistiek en Cultur*) — que tout n'est pas compris dans les nombres, que les caractères purement descriptifs peuvent être aussi importants que les caractères qualitatifs, si non plus importants, surtout pour se rendre compte des manifestations le plus élevées de l'esprit. Autrement dit, il faut resister à la suggestion monopolisatrice des chiffres pour caractériser choses et personnes, suggestion qui fait prise surtout sur les esprits les plus simples. Ce n'est pas par hasard qu'elle fait prise surtout en Amérique et c'est à raison que nous Européens protestons, par exemple, contre l'habitude que l'on y a d'évaluer une personne d'après son revenu.

Il faut dire d'autre part que la manie de tout chiffrer, dont je parlais en ouvrant mon discours et qui est particulièrement prononcée aux Etats Unis d'Amérique, porte en elle même le remède à l'inconvénient que nous venons de signaler, car elle porte à inventer ou à perfectionner de plus en plus les méthodes

pour mesurer des caractères que par le passé on ne savait que décrire : tels l'intelligence, au moyen de « tests » mentaux, le progrès, la solidarité sociale, etc.

Le pouvoir suggestif des chiffres offre dans quelques domaines, des manifestations particulières qu'il vaut la peine d'étudier de près.

Il est en effet des domaines, où les nombres et leur usage provoquent chez l'homme un état de quasi hypnose, tant et si bien qu'il est rationnellement incapable de se soustraire à eux.

Nous en avons eu un témoignage dans le déplacement de l'heure. Lorsqu'on proposa d'avancer les horloges et d'établir une heure d'été qui décalât de 60 minutes toute la vie sociale, en vue d'épargner le courant électrique, tous ceux — et c'étaient la très grande majorité — qui jugeaient l'homme un être raisonnable, se mirent à rire. N'était-ce point ridicule de croire qu'un être raisonnable réglât sa vie suivant les aiguilles d'une montre plutôt que suivant les exigences de la lumière du soleil ? Mais les faits prouvèrent que, effectivement, l'homme est un être d'habitudes plus que de raison, aussi dans ce domaine. Une vaste enquête que j'ai menée alors, permettait de calculer — avec toutes les réserves dues aux difficultés des constatations — que l'économie de courant d'éclairage ainsi obtenue quotidiennement était d'environ 5 à 15%. Presque tous les Etats ont donc adopté l'heure d'été — sauf l'Italie, comme si elle était trop riche pour que ses dirigeants daignent s'occuper de telles vétilles.

Un autre exemple du pouvoir hypnotique que les nombres exercent sur notre esprit nous est donné par les appointements, qui numériquement, restent les mêmes, mais auxquels il faut ajouter, à la fin de l'année, une 13ème mensualité et, dans des nombreuses administrations, une 14ème et même une quinzième. Si l'on réfléchit, on se rend compte que c'est exactement comme si les appointements mensuels subissaient une augmentation de 1, 2 ou 3 douzièmes.

Mais les conséquences peut-être les plus importantes de ce pouvoir hypnotique sont celles qu'on peut constater alors que les prix augmentent ou diminuent.

Il suffit de réfléchir pour comprendre que, si tous les prix varient dans la même mesure, c'est comme s'ils ne variaient pas du tout ; mais lorsque les prix des denrées alimentaires augmen-

tent, les ménagères se désespèrent sur la cherté de la vie, sans tenir compte de ce que les recettes familiales ont augmenté, peut-être même en plus grande proportion. Même un mari économiste, qui sait théoriquement parfaitement bien que sa femme a tort, a toutes les peines du monde à dominer son mouvement instictif de solidarité à l'égard de cette réaction pessimiste.

Par contre, lorsque les prix diminuent, les consommateurs soupirent de soulagement, sans s'inquiéter du fait que leurs revenus aussi sont destinés à diminuer proportionnellement, ou même davantage. Des illusions analogues règlent la conduite des débiteurs et des créditeurs, qui, hypnotisés par le nombre, ne se rendent pas exactement compte de ce qu'ils perdent ou gagnent par suite d'une variation de prix.

Les conséquences de ces illusions sont de la plus grande importance pour la dynamique économique.

Lorsque les prix augmentent, l'illusion de la vie chère pousse à l'épargne ; lorsque les prix baissent, l'illusion de la vie à bon marché pousse à la consommation.

Du moment que l'augmentation des prix a lieu lorsque la demande dépasse l'offre, et la baisse lorsque c'est, au contraire, l'offre qui dépasse la demande, ces illusions hâtent le rétablissement de l'équilibre entre la demande et l'offre, prévenant ainsi, parfois, une crise de déficit ou d'excédent, ou la limitant lorsqu'elle a déjà éclaté.

Lorsque les prix montent, les bilans monétaires des producteurs se ferment sur un actif, même si les bilans réels sont effectivement équilibrés ou, en une certaine mesure, déficitaires. L'activité productive en est stimulée. Elle l'est d'autant plus que les épargnants ne se rendent pas compte de combien ils perdent en accordant des prêts alors que les prix montent, puisque l'intérêt plus élevé ne suffit généralement même pas à compenser la dévaluation effective du capital. Tout aussi naturellement, les débiteurs — qui sont généralement producteurs — gagnent. Ce sont là encore autant de moyens par lesquels l'équilibre entre l'offre et la demande, se rétablit et les crises de déficit, sinon prévenues, en restent allégées.

Toutefois, il s'ensuit également la probabilité que la production, par force d'inertie, dépasse la limite raisonnable, et que l'on passe ainsi de la phase croissante des prix à la phase décroissante.

Lorsque les prix baissent, le bilan monétaire du producteur tend à se clore sur un déficit, même si le bilan réel est équilibré ou actif, si bien que l'on hésite à produire encore ; mais, par ailleurs, les producteurs ne se rendent pas bien compte de la gravité progressive de leur dette, par suite du réajustement de l'unité monétaire — dette qui n'est compensée qu'en partie par la diminution du taux d'intérêt, — si bien que la production, quoique découragée, ne s'arrête jamais entièrement.

Il s'ensuit tout un ensemble de phénomènes dans la dynamique économique, avec des retours périodiques d'expansion ou de contraction, celle-ci prenant parfois la gravité d'une crise. Leur portée et leur étiologie forment l'objet incessant des recherches des économistes.

Si, lors des phases croissantes de prix, on reconnaît généralement opportun que la production en soit stimulée et la consommation limitée, il y a lieu au contraire de se demander à quoi servent les gains extraordinaires obtenus par les commerçants grâce à la différence entre le prix de revient et le prix de vente. Et, en effet, l'opinion publique s'y rebelle, parfois même par des réactions excessives. Ici encore, l'hypnotisme des chiffres a des effets bienfaisants, car, grâce à lui, le vendeur se règle, en établissant le prix, non seulement et peut-être moins sur le prix de reproduction que sur le coût de production (à savoir sur le prix d'achat) limitant ainsi ses gains et freinant, du même coup, la montée des prix. Seulement lorsque leur augmentation est trop violente, le choc qu'en reçoit l'esprit humain est tel qu'il dissipe l'hypnotisme du nombre, et le commerçant règle alors ses prix, non plus sur le coût de production, mais sur celui de reproduction.

Ces périodes exceptionnelles mises à part, une plus grande vulgarisation des connaissances économiques corrige aujourd'hui ces illusions, et les variations des salaires et des appointements commencent à être, en partie, réglées d'après les indices du pouvoir d'achat de la monnaie ; mais en partie seulement, car leur application n'a pas encore été étendue aux rapports entre débiteurs et créditeurs.

Une autre question d'actualité, où l'on constate une fois de plus la force hypnotique du nombre, est celle de la limite d'âge pour la mise à la retraite des fonctionnaires.

Je vais vous entretenir brièvement sur ce sujet, étant donné qu'il présente un intérêt indéniable, de nombreux points de vue, pour le statisticien.

La question des vieillards est à l'ordre du jour. Leur pourcentage, par rapport à la population totale, s'est accru et continue à s'accroître. Dès à présent, leur entretien n'est pas chose facile. On en vient à se demander ce qu'il en sera demain. En effet, ce n'est pas seulement leur nombre proportionnel qui augmente, mais aussi leur influence ; et, pour ces deux raisons, l'intérêt qu'on leur porte et les mesures prises à leur égard s'accroissent également. Des congrès, des sociétés, des revues sont consacrés à l'étude de ces problèmes. La « gérontologie » est devenue une nouvelle science. Il faut s'attendre donc à ce que les frais de leur entretien augmentent plus que proportionnellement, par rapport à leur nombre. Pour y faire face, il faudrait que leurs revenus aussi — qui sont aujourd'hui négligeables — s'accroissent. Cela serait possible et eux-mêmes le souhaiteraient. Cela serait possible, non seulement parce que, pour certaines tâches du moins, leurs facultés pourraient être mieux exploitées qu'elles ne le sont à présent, mais aussi parce que les facultés des personnes âgées s'améliorent progressivement, comme résultat du prolongement de la vie. Alors qu'à un certain âge on avait atteint, hier, la décrépitude, on commence à peine à être vieux maintenant et l'on est encore dans la plénitude de l'âge, alors qu'hier on était déjà un vieillard.

Mais pourquoi donc ne pas retarder la limite d'âge pour la retraite ? Cela satisfairait les personnes âgées, et, tout en allégeant leur existence, allégerait leur fardeau pour la communauté. C'est une solution que l'on demande, mais à laquelle on n'a pas recours justement parce que l'esprit humain est hypnotisé par les nombres. La vieillesse commence officiellement à un certain âge, et l'on n'ose pas en déplacer la limite.

Tel est le tableau qui se présente à nous. En réalité, tout en restant fidèles à ses grandes lignes, nous devrions le retoucher.

Il faut, tout d'abord, rappeler une des plus vieilles régularités démographiques, connue sous le nom de loi de Sundbaerg. Selon cette loi, formulée il y a une cinquantaine d'années et suffisamment vérifiée, les vivants de 15 à 50 ans, chez les diverses populations, représentent un pourcentage à peu près constant de 50% ; en d'autres termes, la vie économiquement pro-

ductive (ou plus productive) représente, dans l'ensemble de la vie humaine, une proportion constante, et cela aussi bien pour les populations stationnaires que pour les populations croissantes ou décroissantes, sans compter les effets des mouvements migratoires. Si, au lieu de la limite de 50, on choisit celle de 60 ans, la proportion, naturellement, change, mais la constante subsiste, avec une bonne approximation, pour les populations stationnaires, et, avec une approximation plus vague, pour les populations en phase croissante ou décroissante.

C'est là un résultat important, pour la question qui nous intéresse. Si nous admettons la validité de la loi de Sundbaerg, nous n'avons pas à nous inquiéter du pourcentage croissant des vieillards, qui serait en grande partie compensé par la diminution du pourcentage des jeunes qui n'ont pas encore atteint la maturité.

Mais, sa validité est subordonnée à une hypothèse que l'on ne doit pas perdre de vue : l'hypothèse que, même si la population n'est pas stationnaire, les coefficients spécifiques de natalité et de mortalité pour chaque âge restent constants. De cette manière, la population s'avère continuellement croissante ou continuellement décroissante, et tend (après un certain temps elle y parvient) à une composition, par âge, qui s'avère constante [1] : nous nous trouvons alors en présence de ce qu'on appelle une *population stable*.

Or la réalité s'écarte de cette hypothèse à deux points de vue.

D'abord, la mortalité spécifique pour chaque âge ne reste pas constante ; au contraire, ainsi que nous l'avons dit, la vie s'allonge, c'est-à-dire que la mortalité spécifique se réduit.

Deuxièmement, la natalité spécifique pour chaque âge ne reste constante non plus, mais présente des variations cycliques importantes.

C'est surtout à cause de ces cycles, et de la dynamique des naissances qui s'ensuit, que la population européenne, au cours des années écoulées, n'était pas stable ; elle ne le sera pas, non

---

[1] La composition par âge reste constante même si les coefficents spécifiques de mortalité pour chaque âge présentent des caractères uniformes (c'est-a-dire de la même intensité absolue) pourvu que les coefficients spécifiques de natalité restent constants.

plus, dans les années à venir. En effet pendant les dix premières années de notre siècle, la dynamique des naissances est arrivée à un tournant, passant de la phase croissante à la phase décroissante. En quelques dizaines d'années, le pourcentage des couches de population en âge de produire est ainsi devenu anormalement élevé, par rapport aux couches plus jeunes et plus âgées. C'est là une condition dans laquelle une grande partie des nations européennes se sont trouvées alors et se trouvent encore aujourd'hui. C'est une condition favorable, non seulement du point de vue économique, mais aussi du point de vue démographique, parce que les hommes ayant de 15 à 50 ou 60 ans, outre qu'ils sont pratiquement les seuls actifs du point de vue économique, sont aussi les seuls, pratiquement, qui peuvent pourvoir à la reproduction démographique et, en même temps, présentent une mortalité bien inférieure à celle de la masse générale de la population. D'où euphorie économique et inflation démographique.

Vers 1935, nous assistons à un autre tournant, chez la plupart des populations blanches. Contre toute attente, les naissances, qui diminuaient auparavant, ont recommencé à augmenter sensiblement ; la courbe ascendante n'a même pas subi de ralentissement pendant et après la guerre. Continuera-t-elle ?

De toute manière, elle a continué jusqu'à présent, et cela suffit, pour que, dans quelques années, les couches adultes de la population, économiquement productives et démographiquement reproductives, diminuent sensiblement auprès des couches des jeunes ou des vieillards qu'elles doivent entretenir. Les perspectives ne sont pas très encourageantes.

Mais n'est-ce point là une raison pour mieux exploiter les facultés des personnes âgées ? Pour proroger la limite d'âge de la retraite, en se soustrayant à l'hypnotisme des chiffres ? Ne serait-ce que provisoirement, tant que dure la période démographiquement et économiquement défavorable, tout comme on adopte provisoirement l'heure d'été ? Ou, mieux encore, suivant un système souple, analogue à celui par lequel, grâce aux nombres indices, on tient compte des variations du pouvoir d'achat de la monnaie ?

Par ailleurs, il ne convient point d'exagérer la vitalité actuelle des vieillards, par rapport au passé, en se basant sur l'accroissement de leur pourcentage. Ce que je viens de dire

laisse comprendre, en vérité, que cette dernière peut dépendre essentiellement de la dynamique passée du nombre de naissances.

La durée moyenne accrue de la vie des populations n'est pas non plus un indice certain de la plus grande vitalité des vieillards : loin de là ! Elle peut dépendre — et pendant un certain temps, elle a dépendu exclusivement — de la diminution de la mortalité chez les enfants. Mais ne pouvons-nous pas établir un indice de vitalité des vieillards moins incertain que celui de l'âge tel que nous l'indique le calendrier ? La réponse est affirmative. Nous pouvons, en effet, calculer cet indice d'après leur vie future présumée. Supposons, par exemple, qu'à une certaine période, — disons 1930 — que nous choisirons comme base — l'âge limite pour la retraite ait été de 70 ans ; pendant les années successives, il pourra être fixé à l'âge auquel correspond une vie moyenne ou une vie probable égale à celle qui, en 1930 correspondait à 70 ans. Cela suppose que la vitalité d'une catégorie d'individus du même âge se mesure d'après l'intervalle absolu qui, en moyenne, les sépare de la mort. D'autres pourraient préferer l'intervalle relatif, à savoir proportionnel à la longueur de la vie passée. Que l'on choisisse l'un ou l'autre, le principe reste.

Construire d'année en année un indice de la vitalité des vieillards est parfaitement possible, mais il serait peut-être souhaitable d'établir des périodes plus longues, afin de compenser les variations accidentelles, par rapport à la tendance générale — variations dues, moins au nombre restreint des cas qu'aux circonstances transitoires de climat, d'hygiène, d'économie. Il vaudrait mieux encore déterminer, sur la base des indices passés, la variation que l'on peut escompter l'année suivante, tout en la rectifiant d'année en année.

Il serait également souhaitable de calculer, auprès de l'indice de la vitalité pour l'ensemble des vieillards, des indices pour certaines de leurs catégories, tout comme, ne se contentant pas d'un indice général des prix, on évalue aussi des indices pour certaines catégories de biens ou pour certaines régions.

Je crois que cela ouvre un terrain d'enquête intéressant ; enquête dont les résultats peuvent nous aider à surmonter la crise qui se prépare, due à un excès de vieillards.

Il est certain de toutes façons, que, dans ce domaine comme dans les autres, il ne faut pas se laisser hypnotiser par les nom-

bres. Si nous devons nous incliner devant la nécessité que la vie moderne soit de plus en plus strictement — et même tyranniquement, avouons-le — reglée par les chiffres pour répondre aux exigences rigoreuses d'une mécanisme social de plus en plus compliqué, nous devons toutefois veiller à ce que le gouvernement des chiffres ne se cristallise point, mais évolue, en répondant, par des adaptations appropriées, à l'évolution individuelle et sociale.

CORRADO GINI

Di alcuni simboli che può essere opportuno impiegare nella trattazione dei fenomeni statistici.

La graduale introduzione di appositi simboli ha immensamente facilitato la trattazione dei problemi teorici delle matematiche e le applicazioni di queste. Rappresenta, codesto, un processo che evidentemente non è ancora ultimato. Nell'esame di alcuni problemi di statistica metodologica e applicata, mi è accaduto invero più volte di imbattermi in equivoci che si sarebbero potuti evitare mediante l'uso di simboli nuovi o anche di simboli già proposti, ma non di uso corrente.

Richiamati tali problemi, tratterò in questa nota di alcuni nuovi simboli che mi pare possano presentare qualche utilità. Non intendo che essi debbano essere usati sempre, ma intendo che essi possano essere opportunamente usati in talune questioni in cui il non usarli può lasciare luogo ad equivoci.

L'argomento fu già delibato in varie occasioni: anzitutto in una comunicazione fatta nel 1946 al Seminario dell'Istituto di Statistica dell'Università di Roma [1], e poi in un corso di lezioni sulle applicazioni dei metodi statistici alle scienze biologiche, svolto nel marzo 1949 all'Università di Barcellona [2] e, nello stesso anno, in una conferenza tenuta all'Università di Madrid su invito di quell'Istituto di Statistica. Prima di dare

---

[1] Seduta del 9 luglio 1946. Vedi Atti del detto Seminario per gli anni accademici 1945/46, 1946/47, 1947-48 in appendice agli « Atti della VIII Riunione Scientifica, Roma, 1-2 Giugno 1949 » della Società It. di Statistica, Roma, 1951, pagg. 205 e 248.

[2] Un riassunto delle dieci lezioni, di cui il corso si componeva, è stato riprodotto al ciclostilo dal Dr. A. Prevosti sotto il titolo *Algunas cuestiones fundamentales para la aplicación de los metodos estadisticos a las ciencias biologicas*, Laboratorio de Antropologia de la Universidad de Barcelona, 1950.

le mie proposte alle stampe, mi è parso opportuno di presentarle ad un uditorio più largo e a carattere internazionale, quale è questo dell'Istituto Inter. di Statistica, lieto se esse potranno dar luogo a discussioni e a suggerimenti di cui possa tener conto (1).

* * *

Nella trattazione delle questioni monetarie ricorre spesso la cosiddetta « equazione dello scambio » rappresentata dalla formula

$$M \cdot V = P \cdot Q \qquad (1)$$

in cui M indica la massa della moneta (o dei suoi surrogati), V la velocità media di una sua unità, Q il valore delle merci scambiate a mezzo della moneta (o dei suoi surrogati) e P il loro prezzo medio.

Ricerche laboriose furono eseguite da vari economisti, talvolta di grido, per « verificare », come essi dicono, l'esattezza di tale uguaglianza (2). Senonchè tale verifica non ha senso, in quanto non si tratta di un' « equazione », ma di un'« identità », che è necessariamente valida.

In matematica si distinguono, invero, tra le uguaglianze, le « identità » che sono valide qualunque siano i valori attribuiti ai termini incogniti che in esse ricorrono, dalle « equazioni » che risultano valide solo per particolari valori di detti termini, valori che si tratta appunto di determinare.

Ora, nel caso della cosidetta equazione dello scambio, la somma dei valori delle merci scambiate in una unità di tempo, può esprimersi in due modi. Può esprimersi mediante la somma delle unità monetarie esistenti, ciascuna moltiplicata per il numero di volte in cui è stata impiegata negli scambi: indicheremo tale somma con $\sum_{i=1}^{n} v_i m_i$, dove m_i indica l'unità monetaria i^{ma}

(1) Questa memoria venne presentata alla XXVII Sessione dell'Istituto Internazionale di Statistica che ebbe luogo a New-Delhi-Calcutta dal 5 al 18 dicembre 1951. Essa viene pubblicata in inglese negli atti di detta Sessione.

(2) Le più note sono quelle del prof. E.W. Kemmerer, *Money and Credit Instruments in their Relation to General Prices*, New York, Holt, 1909 e del prof. I. Fisher, *The Purchasing Power of Money*, New York, MacMillan, 1911.

e v_i la rispettiva velocità di circolazione (vale a dire il numero di volte con cui l'unità monetaria considerata viene usata nell'unità di tempo in parola). Oppure può esprimersi mediante la somma delle merci scambiate, ciascuna moltiplicata per il suo prezzo, somma che indicheremo con $\sum_{c=1}^{s} p_c q_c$, dove q_c indica la merce scambiata c^{ma} e p_c il rispettivo prezzo.

Posto

$$M = \sum_{i=1}^{n} m_i \qquad Q = \sum_{c=1}^{s} q_c$$

$$V = \frac{\sum_{i=1}^{n} v_i m_i}{\sum_{i=1}^{n} m_i} \qquad P = \frac{\sum_{c=1}^{s} p_c q_c}{\sum_{c=1}^{s} q_c}$$

si ottiene l'identità (1), impropriamente chiamata equazione dello scambio.

Verificare la (1) ha lo stesso senso che avrebbe il verificare che 6 più 4 fanno 10. Verificare invero che $6 + 4 = 10$ non avrebbe altro significato che di mostrare che si sanno contare bene 4, 6, 10 oggetti, la validità dell'uguaglianza derivando dalla stessa definizione di 4, 6, 10 ed essendo assolutamente inattaccabile. Similmente è inattaccabile la validità della (1), e le cosiddette verifiche che se ne sono fatte altro significato non hanno se non di dimostrare che le determinazioni, che si sono eseguite delle quantità M, V, Q, P, sono approssimativamente esatte o per lo meno quadrano tra loro in modo soddisfacente [1].

Si tratta di un equivoco — potrà dirsi — in cui sono caduti gli economisti o alcuni economisti, dai quali non si può pretendere una profonda cognizione delle matematiche. Se non che — a parte il fatto che l'esempio forse più noto di tali verifiche è stato dato da Irving Fisher, il quale non solo fu uno dei maggiori economisti contemporanei, ma anche uno dei più versati nelle matematiche — sta di fatto che un'analoga iden-

---

[1] Vedi già, per la sostanza di queste osservazioni, la nostra nota *L'equazione dello scambio e il potere di acquisto della moneta*, in « Rivista Italiana di Sociologia », Roma, 1913.

tificazione di un'identità con un'equazione si incontra nella trattazione della teoria delle proporzioni, collegata ai più bei nomi della matematica greca.

È a Pitagora, invero, che si attribuisce l'introduzione delle tre proporzioni aritmetica, geometrica ed armonica (allora detta subcontraria), che, scritte nella forma algebrica moderna, danno luogo rispettivamente alle formule

$$\frac{a-b}{b-c} = \frac{a}{a}$$

$$\frac{a-b}{b-c} = \frac{a}{b}$$

$$\frac{a-b}{b-c} = \frac{a}{c}$$

ed è lo stesso Pitagora che, accanto a tali proporzioni, collocava la cosidetta « proporzione musicale » data dall'espressione [1]

$$a : \frac{a+b}{2} = \frac{2ab}{a+b} : b$$

Senonchè quest'ultima uguaglianza è un'identità ed effettivamente risulta vera (salvo per $b = 0$ e per $a + b = 0$) qualunque siano i valori attribuiti ad a e b, mentre le tre formule precedenti danno delle equazioni, che risultano vere solo per particolari valori di a, b, c [2].

L'uso di simboli diversi per le equazioni e le identità fu già per vero prospettato. Si è proposto infatti di mantenere, per

---

[1] Vedi, in proposito, Nicomaco di Gerasa, *Introductio aritmetica*, Edizione a cura di R. Hoche, Lipsia, 1866 e Giamblico di Calcide, *De Vita Pythagorica liber*, Edizione Nauck, Pietroburgo, 1844.

[2] Vedi, in proposito, il nostro articolo *L'evoluzione del concetto di media*, in « Metron », Vol. XVI, n. 3-4, 31-IX-1952, che è destinato a costituire un capitolo del trattato su *Le medie*, in preparazione da parecchi anni e già annunciato fin dal 1938 dal Dott. G. Barbensi, nella comunicazione *Cenno su una trattazione monografica delle medie statistiche* fatta alla XXVII Riunione della Società Italiana per il Progresso delle Scienze (Bologna, 4-11 Settembre 1938).

le prime, il consueto segno $=$, e di adottare, invece, per le seconde, il simbolo $\equiv$: in questo campo, si tratta solo di raccomandare, ogni qualvolta sia opportuna, l'adozione di tale proposta, integrandola al fine di distinguere, come diremo, tipi diversi di identità o di equazioni.

Ma, in altri campi delle matematiche — da una parte, nella metodologia statistica e, dall'altra, nel calcolo delle probabilità — l'introduzione di nuovi simboli appropriati sarebbe raccomandabile per evitare, o quanto meno rendere difficili, equivoci della cui portata solo recentemente ci si è resi pieno conto.

La differenza fra la probabilità p di un evento e la frequenza f con cui esso si è verificato è, salvo il segno, evidentemente uguale alla differenza tra la frequenza f con cui l'evento si è verificato e la probabilità p che gli compete ; senonchè, la probabilità che si verifichi tale differenza non è punto la stessa quando sia noto p ed ignoto f o, invece, quando sia noto f ed ignoto p.

La mancanza di simboli differenziati per indicare le due probabilità ha contribuito certamente a nascondere l'infondatezza di una pretesa dimostrazione dell'inversione del teorema di Bernoulli, che, già enunciata, accanto ad una dimostrazione rigorosa e debitamente circostanziata, da Laplace e Poisson [1], era stata in seguito messa da parte, ma ha avuto l'onore immeritato di venire ripresentata in moderni trattati di calcolo delle probabilità, da quello del Bachelier a quello dello Czuber [2].

Analoga questione si presenta a proposito delle relazioni tra le costanti di una massa o universo e quelle dei campioni. Al valore di una certa costante, che caratterizza una massa od universo, fanno riscontro i molteplici valori delle costanti dei

---

[1] Vedi P.S. Laplace, *Théorie analythique des probabilités*, 2ème édition, Paris, 1814, pag. 281-282, S.D. Poisson, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités*, Paris, Bachelier, 1837, pagg. 209-210 e 210-213.

[2] L. Bachelier, *Calcul des probabilités*, Tome I, Paris, Gauthier-Villars, 1912, pagg. 471-472 ; E. Czuber, *Wahrscheinlichkeitsrechnung*, Teubner, Leipzig, Dritte Auflage, 1914 : Vierte Auflage, 1924. Vedi, per tale questione, il nostro discorso inaugurale su *I testi di significatività*, alla VII riunione della Società Italiana di Statistica, Giugno-Luglio 1943.

campioni che se ne possono estrarre e, mediante il calcolo delle probabilità, si può dedurre, dal valore di quella, i valori di questi nonchè le probabilità che ad ognuno di questi compete. Da ciò stesso discende che, dal complesso dei valori delle costanti dei campioni, si può risalire al valore della costante della massa od universo; ma ciò non significa, come da molti moderni statistici, confondendo, si pretende, che dal valore della costante in un singolo campione si possa risalire al valore della costante della massa od universo (1).

Anche tale confusione è stata certamente facilitata dalla mancanza di simboli adatti.

Se questi possono essere riguardati come seri inconvenienti nel campo teorico, non meno grave è stato l'inconveniente, a cui, nel campo pratico, ha dato luogo il non possedere simboli differenziati per indicare la relazione che intercede tra la misura di un carattere e la misura media del carattere corrispondente (per esempio, tra il peso di una persona e il rispettivo perimetro toracico medio) e la relazione che intercede tra la misura del secondo carattere e la misura media del primo. Se ad una statura di m. 1,60 corrisponde, ad esempio, nella media dei casi, un perimetro toracico di m. 0,80, non è detto affatto che ad un perimetro toracico di m. 0,80 corrisponda una statura media di m. 1,60, e il non essersi resi conto di questa differenza ha portato ad adottare, nelle visite delle reclute, un criterio che conduce, come ho mostrato, a riguardare come normali, nel 20% circa dei casi, individui di statura alta che in realtà avrebbero dovuto essere giudicati di perimetro toracico ristretto o, viceversa, a riguardare come deficienti di torace individui bassi di statura che avrebbero dovuto essere considerati come normali (2).

---

(1) Vedi, in proposito, il nostro discorso su *I testi di significatività*, con cui fu inaugurata, il 26 giugno 1943, la VII Riunione scientifica della Società Italiana di Statistica, pubblicato in « Atti della VI e VII Riunione scientifica, Roma, Gennaio 1943 e Giugno 1943 », Roma, 1945, e l'articolo *Statistical Relations and their Inversions*, più innanzi citato alla nota (2) a pag. 24.

(2) Vedi, in proposito, la nostra conferenza tenuta alla Facoltà di Medicina dell'Università di Bruxelles il 3 giugno 1939 su *Les bases de la Biotypologie et les criteriums du récrutement militaire*, pubblicata nella « Revue anthropologique », Juillet-Septembre 1939, e, per maggiori particolari,

* * *

Abbiamo detto che, tra le uguaglianze, si distinguono in matematica le *identità*, che sono valide qualunque siano i valori dei termini incogniti che in essi ricorrono e potrebbero anche dirsi *uguaglianze assolute*, dalle *equazioni*, che sono valide solo per particolari valori di termini incogniti e potrebbero anche dirsi *uguaglianze relative*.

Le identità, a cui ci riferivamo, si possono dire *identità generiche*, in quanto si istituiscono fra classi di valori, di cui resta incognito il valore specifico nei singoli casi, e sono *identità quantitative*. Ma, oltre alle identità quantitative, vi sono *identità qualitative* e, oltre alle identità generiche, vi sono *identità*, che diremo *specifiche*, le quali vengono istituite tra casi singoli. Per esempio, le identità $2 \equiv 1 + 1$; 1 Kg. $\equiv$ 1.000 gr., sono identità specifiche quantitative ; le identità Cesare Borgia $\equiv$ Duca Valentino ; Eroe dei due mondi $\equiv$ Giuseppe Garibaldi ; Rastignac $\equiv$ Vincenzo Morello ; per alcuni storici, Shakespeare $\equiv$ Bacone, sono identità specifiche qualitative.

Nel seguito di questo articolo, parlando di identità senza ulteriore precisazione, intendiamo riferirci sempre alle identità generiche.

Sia per le identità che per le equazioni, considereremo unicamente valori finiti.

Tra le identità così intese, come tra le equazioni, si possono opportunamente fare ulteriori distinzioni.

Vi sono identità, i cui termini incogniti possono assumere valori quali si siano, senza alcuna limitazione, ed altre i cui termini incogniti non possono assumere determinati valori o valori che rientrano in determinate categorie, per esempio valori positivi. Chiameremo le prime *identità illimitate* e le seconde *identità limitate*. Tra le identità limitate, chiameremo *quasi illimitate* le identità che possono assumere qualunque valore, salvo uno od alcuni valori, sempre in numero finito. Riservando il simbolo già in uso $\equiv$ per indicare genericamente qualunque identità,

---

le due note nella « Revue de l'Institut Int. de Statistique » (1937, n. 2 e 1937, n. 3) dal titolo *Une question importante pour la science des constitutions et pour la médecine militaire : comment juger si les proportions d'un individu sont normales ?*

proponiamo di adottare, per le identità illimitate, il simbolo $\equiv$ e, per le identità quasi illimitate, il simbolo $\doteq$, per le identità limitate, il simbolo $\leqq$. Così

$$a + a \equiv 2a \tag{2}$$

$$a \cdot a \equiv a^2 \tag{3}$$

sono identità illimitate che risultano vere quali si siano i valori di a, mentre

$$\frac{a}{b} \cdot b \doteq a \tag{4}$$

$$a : \frac{a+b}{2} \doteq \frac{2ab}{a+b} : b \tag{5}$$

(di cui l'ultima è la già ricordata proporzione musicale dei pitagorici) sono identità quasi illimitate, che risultano vere quali si siano i valori che assumono a e b, salvo per $b = 0$ nella (4) e per $b = 0$ o $(a + b) = 0$ nella (5). Infine

$$a + a \leqq 2|a| \tag{6}$$

$$|a| \cdot a \leqq -a^2 \tag{7}$$

sono identità limitate, di cui la prima risulta valida per qualunque a positivo e la seconda per qualunque a negativo.

Analogamente, sono due identità illimitate la uguaglianza $M \cdot V \equiv P \cdot Q$, erroneamente qualificata «equazione dello scambio» e l'uguaglianza $n + i - m - e \equiv a$ (in cui n significa «nati», i «immigrati», m «morti» e «emigrati», a «incremento della popolazione»), di cui, conosciuti a, n ed m, ci si serve in demografia per determinare l'emigrazione netta $e - i$.

Si avverta che il dire che, in una identità illimitata, i termini incogniti possono assumere valori quali si siano, non significa che essi siano tra loro indipendenti, cosicchè, dato uno di essi, gli altri possano assumere indipendentemente da questo, valori qualsiasi. Basta che possa assumere valori qualsiasi il termine incognito che viene considerato come variabile indipendente. Così, nella identità dello scambio, il livello dei prezzi P

non è punto indipendente dalla massa di moneta M, ma l'uguaglianza rimane valida qualunque sia l'ammontare della massa monetaria o il livello dei prezzi, considerato indipendente dalle altre quantità.

Parleremo infine di *identità determinata* quando, in un certo caso, il termine incognito assume una modalità determinata, come quando l'autore di un furto (per esempio nella banca K) viene identificato nella persona A. T. o quando la soluzione (unica o plurima) di una equazione (per esempio, $x + x^2 = 90$) è individuata in certi valori (nel caso nostro 9 e — 10).

In tali casi, potrà usarsi, per simbolo dell'identità, il segno $\equiv$, scrivendo, negli esempi su citati,

$$\text{Ladro} \equiv \text{A.T.} \qquad (8)$$

$$x \equiv 9 \text{ ovvero } -10 \qquad (9)$$

* * *

Tra le equazioni, vi ha luogo analogamente a distinguere le *equazioni determinate*, in cui ognuno dei termini incogniti che in esse ricorre può assumere un solo valore o certi determinati valori, e le equazioni *indeterminate*, in cui ognuno dei termini incogniti può assumere valori svariati. Ancora, tra le equazioni indeterminate, si possono distinguere, analogamente a quanto abbiamo fatto per le identità, le *equazioni indeterminate illimitate*, in cui ognuno dei termini incogniti può assumere qualunque valore, i diversi loro valori dovendo però essere tra loro coordinati in modo da soddisfare l'equazione; le *equazioni indeterminate quasi illimitate*, in cui i termini incogniti possono assumere valori qualsiasi, purchè tra loro coordinati, fatta eccezione per alcuni particolari valori, sempre in numero finito, e le *equazioni indeterminate limitate*, in cui i termini incogniti possono assumere solo valori tra loro coordinati che rientrano in determinate categorie, ma non possono assumerne altri che rientrano in categorie diverse.

Riservando il simbolo $=$ per indicare genericamente qualunque uguaglianza, useremo del simbolo $\asymp$ per le equazioni indeterminate illimitate, del simbolo $\pm$ per le equazioni indeterminate quasi illimitate, del simbolo $\simeq$ per le equazioni indeterminate limitate, e, infine, del simbolo $\doteq$ per le equazioni de-

terminate. Così

$$a \equiv -a \qquad (10)$$

$$2a \equiv a \qquad (11)$$

$$\frac{2}{a} \equiv 2 \qquad (12)$$

$$a + a \equiv a^2 \qquad (13)$$

sono equazioni determinate di cui la (10) e la (11) risultano valide solo per il particolare valore $a = 0$, la (12) per $a = 1$ e la (13) per $a = 2$ o $a = 0$;

$$a^3 + b^3 \equiv c^3 \qquad (14)$$

è una equazione indeterminata limitata che non risulta valida per a, b, c interi;

$$\frac{a - b - 2}{a - b^2} \pm c \qquad (14 \text{ bis})$$

è una equazione indeterminata quasi illimitata, in cui a e b possono assumere valori qualunque, tra loro coordinati, purchè non sia $a = 4$ e $b = 2$, nè $a = 1$ e $b = -1$,

$$a + b \equiv c \qquad (15)$$

è una equazione indeterminata illimitata, in cui a, b, c possono assumere valori qualunque, purchè tra essi opportunamente coordinati.

Sarà opportuno indicare, a fianco delle identità o delle equazioni determinate o limitate o quasi illimitate, la categoria di valori o i valori per cui l'identità od equazione è valida, oppure la categoria di valori o i valori per cui non è valida.

Converremo di indicare tra parentesi tonde i valori o la categoria di valori per cui è valida, e invece tra parentesi quadre i valori o la categoria di valori per cui non è valida. Le identità (4) e (5) si possono quindi completare nel modo seguente:

$$\frac{a}{b} \cdot b \equiv a \qquad [b = 0] \qquad (4')$$

$$a : \frac{a + b}{2} \equiv \frac{2ab}{a + b} : b \qquad [b = 0 \,,\, a + b = 0] \qquad (5')$$

e le (6) e (7) nel modo seguente:

$$a + a \mathrel{\overline{\equiv}} 2|a| \qquad (a \geq 0) \qquad (6')$$

$$a \cdot a \mathrel{\overline{\equiv}} -a^2 \qquad (a \leq 0) \qquad (7')$$

oppure

$$a + a \mathrel{\overline{\equiv}} 2|a| \qquad [a < 0] \qquad (6')$$

$$a \cdot a \mathrel{\overline{\equiv}} -a^2 \qquad [a > 0] \qquad (7')$$

e analogamente le identità (8) e (9) come segue:

$$\text{Ladro} \equiv \text{A.T} \qquad (\text{Furto Banca } K) \qquad (8')$$

$$x \equiv \begin{cases} +9 \\ -10 \end{cases} \qquad (x + x^2 = 90) \qquad (9')$$

e le equazioni (10), (11), (12), (13), (14) come segue:

$$a \mathrel{\overline{=}} -a \qquad (a = 0) \qquad (10')$$

$$2a \mathrel{\overline{=}} 0 \qquad (a = 0) \qquad (11')$$

$$\frac{2}{a} \mathrel{\overline{=}} 2 \qquad (a = 1) \qquad (12')$$

$$a + a \mathrel{\overline{=}} a^2 \qquad \left(a = \begin{cases} 0 \\ 2 \end{cases}\right) \qquad (13')$$

$$a^3 + b^3 \mathrel{\overline{=}} c^3 \qquad [a, b, c \text{ interi}] \qquad (14')$$

$$\frac{a - b - 2}{a - b^2} \pm c \quad \begin{bmatrix} a = 4 \text{ e } b = 2 \\ a = 1 \text{ e } b = -1 \end{bmatrix} \qquad (14' \text{ bis})$$

È già in uso, per indicare genericamente un'uguaglianza approssimativa, il simbolo $\simeq$; con analogo espediente potranno indicarsi le identità approssimative illimitate, quasi illimitate, limitate o determinate e le equazioni approssimative determinate, indeterminate limitate e indeterminate quasi illimitate e indeterminate illimitate, adottando rispettivamente i simboli: $\overline{\equiv}$, $\overline{\simeq}$, $\overline{\equiv}$, $\overline{\simeq}$, $\overline{\simeq}$, $\overline{\simeq}$, $\overline{\simeq}$, $\overline{\simeq}$.

* * *

In contrapposto al simbolo $=$ di uguaglianza, è di uso comune il simbolo $\neq$, per designare una disuguaglianza.

Mentre nelle uguaglianze abbiamo contrapposto le uguaglianze assolute o identità, che si verificano in ogni caso, alle uguaglianze relative o equazioni, che si verificano solo per certi valori dei termini incogniti, *le disuguaglianze*, come è facile rendersi conto, *sono sempre assolute*. Effettivamente, se tra certe quantità non si può stabilire nè una identità nè una equazione, vale a dire se non si può stabilire una uguaglianza valida nè per tutti i casi nè per alcuni casi, ciò significa che tra dette quantità sussiste una disuguaglianza in tutti i casi.

Analogamente a quanto si è fatto per le identità, le disuguaglianze si potranno distinguere in *illimitate*, se i termini incogniti potranno assumere valori qualsiasi, *quasi illimitate* se essi potranno assumere tutti i valori, salvo ad alcuni, sempre in numero finito, *limitate* se essi non potranno assumere certe categorie di valori (per esempio, solo valori positivi o solo valori negativi) e *determinate*, se potranno assumere soltanto uno o alcuni valori. Esse potranno indicarsi rispettivamente coi simboli $\not=$, $\not<$, $\not>$, $\not\gtrless$ che sono quelli stessi usati per le identità rispettive attraversati da una sbarra obliqua.

È una disuguaglianza illimitata la seguente

$$2 + a + b \not= a + b \tag{16}$$

Sono disuguaglianze quasi illimitate le seguenti

$$a \not= -a \qquad [a = 0] \tag{17}$$

$$2a \not= a \qquad [a = 0] \tag{18}$$

$$\frac{2}{a} \not= 2 \qquad [a = 1] \tag{19}$$

$$a + a \not= a^2 \qquad \left[a = \begin{cases} 0 \\ 2 \end{cases}\right] \tag{20}$$

Notisi come le disuguaglianze quasi illimitate abbiano un campo di validità complementare a quello delle equazioni determinate [cfr. formule (10), (11), (12), (13)].

Sono disuguaglianze limitate le seguenti

$$a^3 + b^3 \not\equiv c^3 \qquad (a, b, c \text{ interi}) \qquad (21)$$

$$a + a \not\equiv 2|a| \qquad (a < 0) \qquad (22)$$

— che può anche scriversi

$$a + a \not\equiv 2|a| \qquad (a \geqslant 0) \qquad (23)$$

$$|a| \cdot a \not\equiv -a^2 \qquad (a > 0) \qquad (24)$$

— che si può pure scrivere

$$|a| \cdot a \not\equiv -a^2 \qquad [a \gtrless 0] \qquad (25)$$

Si avverta come le disuguaglianze limitate abbiano un campo di validità complementare a quello di identità limitate (cfr. formule (6) e (7)) o di equazioni indeterminate limitate (cfr. formula (14).

Sono infine disuguaglianze determinate le seguenti

$$p_n \not\equiv 2k + 1 \qquad (n = 1) \qquad (26)$$

$$p_n \not\equiv 6k \pm 1 \qquad (n = 1, n = 2) \qquad (27)$$

dove p_n indica *l'n^mo* numero primo (e quindi è $p_1 = 2$, $p_2 = 3$), k è un intero appropriato.

Per le disuguaglianze come per le identità si possono considerare, oltre alle *disuguaglianze generiche*, come sono quelle sopra considerate, *disuguaglianze specifiche*, sia quantitative come la seguente

$$1 \not\equiv 2 \qquad (28)$$

sia qualitative come le seguenti

$$\text{Enrico IV di Francia} \not\equiv \text{Enrico IV Imperatore} \qquad (29)$$

e, per molti storici, come per il gran pubblico,

$$\text{Shakespeare} \not\equiv \text{Bacone} \qquad (30)$$

Per designare una forte disuguaglianza, si potrebbero usare due sbarre oblique invece di una, scrivendo, ad esempio,

$$1 \not\equiv 1 \text{ milione} \qquad (31)$$

* * *

Le disuguaglianze quantitative si possono verificare in due sensi a seconda che il primo membro è superiore o inferiore al secondo. Nei segni $>$ (maggiore di) o $<$ (minore di), di uso corrente, si possono introdurre differenziazioni analoghe a quelle sopra proposte per differenziare i segni di disuguaglianza

Si ottengono così i segni seguenti (i primi 4 per indicare che il primo termine è maggiore del secondo e gli ultimi 4 per indicare che il secondo è maggiore del primo)

$$\succ \gg \geqslant \gtrsim \qquad \prec \ll \leqslant \lesssim$$

da adoperarsi a seconda che la disuguaglianza si verifica per tutti i valori incogniti che ricorrono nelle formule, o per tutti salvo uno o qualcuno, o per tutti i valori che rientrano in certe categorie (per esempio, valori positivi o, invece, valori negativi) o infine per uno o alcuni valori determinati.

Si scriverà ad esempio

$$4\,|a| \succ 2\,|a| \qquad (a \neq 0) \tag{32}$$

$$|a| + |b| \gtrsim |a| \qquad [b = 0] \tag{33}$$

$$4a \succ 2a \qquad (a > 0) \tag{34}$$

$$d(n) \prec 2 \qquad (n = 1) \tag{35}$$

dove $d(n)$ indica il numero dei divisori dell'intero n.

I simboli $\gg$ e $\ll$ sono già in uso per significare « molto maggiore di » e « molto minore di » e possono essere applicati in combinazione con le precedenti differenziazioni.

Nelle disuguaglianze quantitative, non è detto che un termine sia sempre maggiore o sempre minore dell'altro. Potrà darsi che talvolta il primo di essi sia maggiore e tale altra minore del secondo e, in tal caso, che, nella media, il primo risulti maggiore, uguale o minore del secondo.

In generale i simboli seguenti

$$\succsim \quad \asymp \quad \precsim$$

indicheranno che il primo termine è talvolta maggiore e talvolta minore del secondo, ma che nella media il primo risulta maggiore o rispettivamente uguale o minore del secondo.

Se con $\dot{u}$ e $\dot{d}$ si indica la età al matrimonio dello sposo e rispettivamente della sposa di una coppia, si potrà scrivere;

$$\dot{u} \asymp \dot{d}$$

* * *

Verrà spontanea a molti la domanda se vale la pena di mettere in moto tutto questo armamentario di simboli quando le quattro semplicissime notazioni in uso $=$, $\neq$, $>$, $<$ pure rispondono bene nella grande maggioranza dei casi. E mi fu anche contestato che questa differenziazione va contro a uno degli scopi fondamentali della matematica, che, secondo un motto di Poincaré, è di usare nomi o simboli uguali per cose diverse.

Si risponde che cogliere il comune nel diverso, risalire, così, dal particolare al generale, non è solo uno degli scopi della matematica, ma di ogni ricerca scientifica che abbia carattere sintetico. Se non che, accanto a questa funzione, certamente utilissima, della scienza, vi è quella, pure utilissima, di cogliere e contrassegnare e analizzare le differenze tra i fenomeni.

Questo per ciò che concerne il punto di vista teorico; per ciò che concerne il punto di vista pratico, la risposta è già stata data preventivamente, quando fin dall'inizio si è osservato che non si intende che i simboli proposti debbano essere usati sempre, ma solo nella trattazione di certi problemi in cui risultano utili. Quando interessa solo il fatto che una quantità o un'espressione abbia un'intensità uguale o diversa, maggiore o minore, di un'altra, bastano le notazioni consuete. Nella stessa presentazione che precede, dei nuovi simboli differenziati, io stesso ho perciò più di una volta usato e userò i correnti simboli generici (cfr. espressioni tra parentesi nelle formule 5'-7', 9'-13', 17-20, 22-27, 34-35, 46). Ma, quando invece importa anche mettere in luce la maggiore o minore generalità o la portata delle uguaglianze o delle disuguaglianze considerate, allora è bene contrassegnarle con simboli diversi ed è precisamente dai constatati inconvenienti a cui ha dato luogo il non differenziarle che ha preso le mosse questa nostra trattazione.

* * *

Le formule che ricorrono nelle uguaglianze o disuguaglianze che abbiamo passato in rassegna, sono tutte estremamente semplici e appunto in ragione della loro semplicità le abbiamo scelte per i nostri esempi.

Tuttavia, se parecchie delle uguaglianze o disuguaglianze che le contengono risultano a prima vista evidenti, derivando immediatamente dalle definizioni delle quantità o dei simboli che in esse ricorrono, altre lo sono già meno. Le formule poi che si incontrano in pratica nei problemi matematici e nella trattazione dei dati statistici sono in generale molto più complicate, così che le uguaglianze o disuguaglianze che tra esse si stabiliscono hanno bisogno di una dimostrazione, talvolta con non pochi passaggi, per decidere se di una uguaglianza o di una disuguaglianza si tratta e, nel secondo caso, in che senso questa si compie.

Finchè a tale decisione non si sia pervenuti, quale segno si deve usare? Sarebbe ugualmente arbitrario servirsi di un segno qualsiasi, vuoi di uguaglianza che di disuguaglianza.

Su tale inconveniente ha richiamato la mia attenzione il prof. Pompilj. Ad ovviarvi mi pare consigliabile adottare, per esprimere la relazione tra due quantità, un segno generico che potrebbe essere il seguente $\ominus$, destinato ad essere sostituito, a dimostrazione compiuta, dal segno appropriato di uguaglianza o di disuguaglianza.

* * *

Ritorniamo alle equazioni indeterminate.

Se in un'equazione indeterminata con n termini incogniti, vengono determinati i valori di $n - 1$ di questi, vi è un numero finito di valori dell'n^{mo} termine che risolve l'equazione, la quale quindi diviene un'equazione determinata. Il numero degli $n - 1$ termini incogniti, che bisogna determinare per rendere l'equazione determinata, può chiamarsi *ordine della indeterminazione*. La (14) e la (15), ad esempio, sono equazioni indeterminate di secondo ordine.

Il valore dell'n^{mo} termine incognito che risolve l'equazione naturalmente varia in corrispondenza o, come suol dirsi, in dipendenza del variare dei valori degli $n - 1$ altri termini incogniti; esso si dice pertanto *variabile dipendente* o *funzione* degli altri $n - 1$ termini che costituiscono le *variabili indipendenti*. Si intende che esse sono considerate come indipendenti rispetto all'n^{mo} termine incognito; chè nulla esclude che esse siano tra loro legate da vincoli di connessione. Quando l'equazione indeterminata è di primo ordine, il valore del termine incognito si dirà *funzione semplice* del valore dell'altro; quando

l'equazione indeterminata è di ordine superiore, il valore di ogni termine incognito si dirà *funzione multipla* del valore degli altri, e si dirà funzione di 2°, di 3° ... di $n - 1^{mo}$ ordine, a seconda che l'ordine dell'indeterminazione è il 2°, il 3°, l'$n - 1^{mo}$, ossia che le variabili indipendenti da cui esso si fa dipendere sono 2, 3 ... $n - 1$.

Una funzione semplice dicesi *biunivoca* quando ad ogni valore della variabile indipendente corrisponde un diverso valore della funzione ed uno solo ; *univoca* quando ad ogni valore della variabile indipendente corrisponde un solo valore della funzione, ma lo stesso valore della funzione può corrispondere a più di un valore della variabile indipendente ; *plurivoca* quando ad ogni valore della variabile indipendente corrisponde un diverso valore, ma non sempre un solo valore della funzione ; *biplurivoca*, infine, diremo una funzione semplice quando ad un valore della variabile indipendente può corrispondere più di un valore della funzione e, nello stesso tempo ad un valore della funzione può corrispondere più di un valore della variabile. Le equazioni

$$a = b^3 \qquad (36)$$

$$a = b^2 \qquad (37)$$

$$a^2 = b \qquad (38)$$

$$a^2 = b^4 \qquad (39)$$

(dove b è variabile ed a funzione) forniscono esempi di una funzione semplice rispettivamente biunivoca, univoca, plurivoca, biplurivoca.

Le funzioni multiple non sono biunivoche, poichè uno stesso valore della funzione può risultare da combinazioni di diversi valori delle variabili indipendenti, salvo che queste siano tutte l'una funzione biunivoca dell'altra, nel qual caso però la funzione solo apparentemente è multipla, ma in realtà è semplice.

Nelle funzioni biunivoche, la corrispondenza tra i valori della variabile indipendente e della funzione è *reciprocamente invertibile*, nel senso che, se ad un dato valore della variabile indipendente corrisponde un dato valore della funzione, a questo valore della funzione corrisponde sempre il detto valore della variabile indipendente, e a questo valore della variabile indipendente corrisponde sempre il detto valore della funzione : per esempio, in

base alla (36), sapendo che è $b = 2$, possiamo dire che è sempre $a = 8$, e saputo così che è $a = 8$, possiamo dire che è sempre $b = 2$. Viceversa, sapendo che è $a = 8$, possiamo dire che è sempre $b = 2$ e, saputo che $b = 2$, possiamo dire che è sempre $a = 8$.

Ciò non si verifica nelle funzioni univoche plurivoche o biplurivoche. Così, in base alla (37), sapendo che $a = 4$, possiamo dire che b è $= +2$ oppure $= -2$ e, in entrambi i casi, da questo valore si ricava che è $b^2 = a = 4$, vale a dire in questo senso la funzione è invertibile; ma, d'altra parte, sapendo che $b = 2$, possiamo dire che è $a = 4$; ma, saputo che è $a = 4$, non possiamo dire che è $b = 2$, potendo anche essere $b = -2$. E, analogamente in base alla (38). In base alla (39), invece, sapendo che è $b = 2$, e quindi $b^4 = 16$, potremo dire che $a = +4$ ovvero -4 e da questo valore potremo risalire ad $a^2 = 16 = b^4$, che però potrà corrispondere a $b = -2$. Analogamente, sapendo che è $a = 4$, ricaveremo $a^2 = 16 = b^4$ da cui però si potrà ricavare $b^2 = a = +4$ oppure $= -4$.

Le funzioni univoche o plurivoche sono quindi *invertibili solo in un senso* e le funzioni biplurivoche *non* sono *invertibili in alcun senso.*

* * *

Nella teoria matematica, le equazioni e le corrispondenti funzioni si studiano per quantità astratte e le difficoltà che in essa si presentano dipendono dalla difficoltà di risolvere le equazioni considerate. Nelle applicazioni, invece, le equazioni si studiano per quantità concrete, che rappresentano la misura di dati fenomeni o caratteri osservati: alle difficoltà teoriche suddette si aggiungono allora gli ostacoli che derivano dalle difficoltà pratiche di ottenere tali misure. Sia le difficoltà teoriche, che le pratiche, crescono in generale, con la complessità della funzione. Ora, nelle applicazioni, l'intensità dei fenomeni o caratteri considerati, è in generale connessa con l'intensità di molteplici altri fenomeni o caratteri, cosicchè le funzioni rispettive risultano di un ordine di complessità molto elevato. Determinare l'intensità di tutti questi fenomeni o caratteri che rappresentano le variabili indipendenti non è, d'altro lato, praticamente possibile, a parte la difficoltà che vi sarebbe, una volta che tali intensità si fossero determinate, a ricavarne la complessa legge funzio-

nale. Nella maggior parte dei casi, neppure riesce possibile di determinare le intensità di poche variabili indipendenti, o magari di una sola, mantenendo immutate le intensità delle altre e riducendo così la funzione ad un grado minore di complessità o magari ad una funzione semplice. Negli esperimenti si riesce, bensì, a mantenere praticamente costanti alcune variabili indipendenti che si ritiene esercitino la massima influenza sulla funzione, ma di molte altre le variazioni non si possono eliminare e il loro effetto viene assimilato alle conseguenze di errori accidentali di osservazione. Nella osservazione dei fatti quali si svolgono in natura è più difficile, e talvolta anzi impossibile, di eliminare anche l'influenza delle variabili indipendenti di maggiore portata. Per esempio, la dimensione del perimetro toracico di un individuo è connessa con la dimensione della sua statura e, a parità di statura, con la razza, la combinazione morfologica, l'età, le condizioni di salute, il nutrimento, la stagione dell'anno, l'ora del giorno, la fase di espirazione o di inspirazione, il livello a cui la misura viene presa, l'esattezza dello strumento adoperato, la precisione del misuratore. Ora, alcune di queste variabili si possono eliminare e di fatto più o meno esattamente si eliminano, prendendo, ad esempio, la misura all'altezza della quarta costola, durante la pausa respiratoria, ed altre si possono eliminare in parte, considerando, ad esempio, solo individui adulti e scartando i malati; ma le rimanenti — di cui alcune sono pure di primaria importanza — di solito non si eliminano e in parte non si potrebbero praticamente eliminare.

Ne viene che, ad una data misura della statura, non corrisponde, per tutti gli individui misurati, la identica misura del perimetro toracico, ma corrispondono misure varie da individuo a individuo. Ciò avviene indipendentemente dal carattere biunivoco o univoco o plurivoco della relazione funzionale tra perimetro toracico e statura, per il fatto che i dati disponibili effettivamente non ritraggono la relazione funzionale semplice tra perimetro toracico e statura, ma la relazione funzionale complessa tra perimetro toracico e una molteplicità di caratteri indipendenti di cui la statura non è che uno. Nell'ipotesi che le influenze di tutte queste variabili indipendenti, tranne la statura, tendano ad eliminarsi, si sostituisce alla relazione funzionale ignota tra perimetro toracico e statura la relazione tra perimetro toracico medio e statura. Non è il caso di esaminare qui — esame

che ho già fatto altra volta [1] — il fondamento di tale ipotesi. Importa e basta qui stabilire che, pure ammesso che tale ipotesi sia fondata, le relazioni, che si considerano in pratica tra due o più fenomeni o caratteri, sono, indipendentemente dalla natura biunivoca o non biunivoca delle funzioni che si vogliono determinare, evidentemente plurivoche. Tali relazioni si dicono *statistiche* o *stocastiche*. In ipotesi particolari, venne anche studiata teoricamente la distribuzione che, intorno alla loro media, presentano i diversi valori della funzione corrispondenti ad un dato valore della variabile, spiegando perchè detta distribuzione spesso assuma in fatto la forma così detta gaussiana, e, nel caso di frequenze relative, venne determinata, oltre che la forma, anche l'ampiezza teorica di tale distribuzione.

Nelle relazioni statistiche, dunque, non è il caso di domandarsi quale è la relazione che passa tra le misure di due fenomeni o caratteri A e B, ma quale è la relazione che passa tra la misura dell'uno (per esempio A) e la misura media dell'altro (rispettivamente B), oppure con quale frequenza si verifica una certa relazione tra una misura di A e una corrispondente misura di B. D'altra parte, queste relazioni non risultano invertibili; vale a dire non è detto che la relazione trovata tra la misura di A e la misura media di B sia uguale alla relazione che passa tra la misura di B e la misura media di A, o che la frequenza, con cui si verifica una certa relazione tra detta misura di A e una corrispondente misura di B, sia uguale alla frequenza con cui si verifica la relazione inversa tra questa misura di B e la anzidetta misura di A.

Devesi ancora tener presente che vi sono varie medie dei valori della funzione che si possono mettere in corrispondenza coi valori della variabile indipendente, e le relazioni trovate per una media possono non essere valide per un'altra.

È quindi opportuno introdurre simboli adatti per

a) distinguere la funzione dalla variabile;

b) distinguere dalle relazioni tra variabile e funzione le relazioni tra variabile e un valore medio della funzione;

---

[1] Vedi l'articolo *Alle basi del metodo statistico. Il principio della compensazione degli errori accidentali e la legge dei grandi numeri*, in « Metron », vol. XIV, n. 2-3-4, 31-XII-1941.

c) distinguere i più importanti valori medi della funzione ;

d) indicare la frequenza con cui si verifica una certa relazione tra un valore (o classe di valori) della variabile e un corrispondente valore (o classe di valori) della funzione e distinguerla dalla frequenza con cui si verifica la relazione inversa tra un valore (o classi di valori) della funzione e un corrispondente valore (o classe di valori) della variabile : ne segue l'opportunità di studiare la migliore maniera per

e) rappresentare la classe dei valori di una variabile che cadono in un certo intervallo.

Esaminerò successivamente i vari punti.

a) Proposta pratica mi pare quella di indicare la variabile con lettera minuscola e la funzione con lettera maiuscola.

Proposta alternativa è di non differenziare le lettere usate per la variabile indipendente e per la funzione, ma di differenziare il segno di uguaglianza facendo assumere alla lineetta superiore di detto segno l'aspetto di una semi-freccia che indichi la direzione con cui si passa dalla variabile indipendente alla funzione.

L'espressione « bi » funzione di « emme » nel primo modo si esprimerà

$$B = m$$

e nel secondo

$$b \leftharpoondown m$$

Come vedremo, ognuno dei due metodi ha i suoi vantaggi e i suoi svantaggi. Faremo uso per ora del primo metodo.

b) e *c*) Le medie dei valori della funzione possono essere indicate con segni convenzionali posti al di sopra della lettera scelta per indicare la funzione : con un trattino orizzontale si indica spesso la media aritmetica ; con un trattino verticale si potrà indicare la mediana ; col segno $\wedge$ la moda, col segno $\vee$ la antimoda, mentre con un punto si indicherà un valore singolo della funzione.

Se, per esempio, con F si indica la funzione, si avrà

$\bar{F}$ = valore medio,

$\dot{F}$ = valore mediano,

$\hat{F}$ = moda, ossia valore di massima frequenza,

$\check{F}$ = antimoda, o valore di minima frequenza,

$\dot{F}$ = un valore singolo della funzione.

Quindi, se con le lettere « esse » e « pi » si indicano le misure della statura e rispettivamente del perimetro toracico, la uguaglianza

$$\overline{P} = k\, s^a \qquad (40)$$

dice che il perimetro toracico medio cresce col crescere della potenza a^{ma} della statura e la uguaglianza

$$\overline{S} = h\, p^b \qquad (41)$$

dice che la statura media cresce col crescere della potenza b^{ma} del perimetro toracico. Quanto abbiamo detto fa capire che sarebbe erroneo ricavare dalla (40) l'uguaglianza

$$\overline{S} = \left(\frac{p}{k}\right)^{\frac{1}{a}} \qquad (42)$$

come sarebbe erroneo ricavare dalla (41) la uguaglianza

$$\overline{P} = \left(\frac{s}{h}\right)^{\frac{1}{b}} \qquad (43)$$

L'introduzione dei simboli sopra indicati mira appunto ad eliminare il pericolo di simili inversioni non fondate.

Se con p indichiamo la probabilità, nota *a priori* e supposta costante, di un evento e con F la frequenza che dell'evento si è osservata in un esperimento di un certo numero n di osservazioni, la teoria dimostra che è da attendersi

$$\overline{F} = p \qquad (44)$$

$$\hat{F} \simeq p \qquad (45)$$

ma

$$\check{F} \neq p \qquad [p = 0{,}5] \qquad (46)$$

essendo

$$\tilde{F} \gtrless p \quad \text{per} \quad p \gtrless 0{,}5 \qquad (46{,}2)$$

In altre parole, se si potesse fare un numero infinito di esperimenti, tutti di n osservazioni, si troverebbe che la frequenza media aritmetica e la frequenza modale coinciderebbero con la probabilità, mentre la frequenza mediana sarebbe diversa dalla probabilità, salvo nel caso particolare $p = 0{,}5$.

Anche qui la (44) e la (45) non sono invertibili. Non si può, cioè, affermare che, quando in un esperimento di n osservazioni relative ad un evento, questo presenta una frequenza f, la probabilità media del suo verificarsi sia, in un numero infinito di esperimenti consimili

$$\bar{P} = f \qquad (47)$$

nè che la probabilità modale sia

$$\hat{P} = f \qquad (48)$$

Per dare un qualsiasi giudizio sul valore di $\bar{P}$ o di $\hat{P}$ è necessario avere qualche cognizione supplementare. Nell'ipotesi (ipotesi di Laplace) che *a priori* si possa ammettere che tutte le probabilità p dell'evento intervengano con la stessa frequenza, la (48) risulta giustificata, ma non la (47), essendo invece da attendersi

$$\bar{P} = \frac{nf + 1}{n + 2} \qquad (49)$$

Nell'ipotesi (ipotesi di Gini) che *a priori* si possa ammettere che ad ogni valore p della probabilità dell'evento competa una frequenza

$$\omega = \frac{\Gamma(k + h + 2)}{\Gamma(k + 1)\,\Gamma(h + 1)}\, p^k (1 - p)^h$$

è da attendersi

$$\bar{\bar{P}} = \frac{nf + k + 1}{n + k + h + 2} \quad (1) \qquad (50)$$

---

(1) Cfr. *Considerazioni sulle probabilità a posteriori e applicazioni al rapporto dei sessi nelle nascite umane*, Università di Cagliari, 1911, riprodotto in « Metron », vol. XV, 15-X-1949, pag. 145.

Un problema analogo si presenta nelle relazioni tra mese di matrimonio e mese di nascita del primogenito. Conosciuto il mese del matrimonio, si può prevedere, in base alle statistiche di cui si dispone, che il più probabile mese di nascita del primogenito sarà il decimo successivo. Da ciò non segue però che, conosciuto invece il mese della nascita del primogenito, si possa affermare che il mese più frequente del matrimonio è stato il decimo precedente: tale previsione sarebbe giustificata se si potesse ammettere che il numero dei matrimoni fosse lo stesso in tutti i mesi. Ma questa ipotesi notoriamente non si avvera e quindi non è da meravigliarsi se, come fu constatato [1], varia, secondo il mese della nascita, l'intervallo più probabile tra esso e il mese del matrimonio. Questo risulta a volte il decimo precedente, ma a volte invece il nono o l'undicesimo o il dodicesimo; nella media però il mese più probabile risulta il decimo.

Ciò si può esprimere scrivendo

$$\hat{N} = c + 9 \qquad (51)$$

$$\hat{C} \asymp n - 9 \qquad (52)$$

dove con «ci» si indica il mese del matrimonio e con «enne» il mese della nascita.

Si dice in tal caso che l'uguaglianza (51) è *subinvertibile* [2].

Vediamo un altro caso di relazioni subinvertibili.

Comunemente si consiglia all'uomo di scegliere per sposa una donna la cui età superi i vent'anni di una metà dell'intervallo per cui egli li supera ed è interessante constatare che tale

---

[1] C. Gini, *Sulla probabilità inversa nel caso di grandezze a distribuzione costante*, Atti della VI e VII Riunione Scientifica, Roma, Gennaio 1943 e Giugno 1943, Roma, 1945, pagg. 283-299.

[2] Il distinguere le relazioni invertibili dalle subinvertibili è di fondamentale importanza per chiarire delicati problemi della metodologia statistica e in particolare la portata del metodo degli intervalli di confidenza. (Vedi il nostro articolo *Statistical Relations and their Inversions* in «Review of the International Statistical Institute», 1947, Edizione tedesca in «Schweizerische Zeitschrift fur Volkswirtschaft und Statistik» 83 Jahrgang, Heft 6, 1947. Vedi anche la successiva nota nella stessa Rivista, 84 Jahrgang, Heft 4, 1948.

relazione si verifica con buona approssimazione tra le età dello sposo e l'età media della sposa. Indicando con « u » l'età dello sposo e con « di » l'età della sposa, si può scrivere

$$\bar{D} = 10 + 0{,}5\,u \qquad (53)$$

Questa relazione non è invertibile, vale a dire la inversa

$$\bar{U} = 2\,(d - 10) \qquad (54)$$

non vale per i singoli valori di d. La relazione (53) è però subinvertibile, vale a dire la inversa (54) vale per la media dei valori di d e di $\bar{U}$, ciò che, facendo uso del simbolo proposto a pag. 14, si può esprimere scrivendo

$$\bar{U} \asymp 2\,(d - 10)$$

d) Proponiamo di adottare il simbolo $\overset{z}{=}$, dove è $0 \leq z \leq 1$, per significare « uguale a... con una frequenza z » e i simboli $\overset{z}{>}$, $\overset{z}{<}$ per significare « maggiore di » o « minore di... con una frequenza z ».

Per esempio, se con n si indica il numero dei figli nati da una coppia e con S il numero dei figli sopravviventi, l'espressione

$$S \overset{60}{>} 0{,}6\,n \qquad (55)$$

significherà che in un terzo delle famiglie il numero dei sopravviventi è maggiore del 60% dei nati.

e) L'espressione della frequenza con cui si verifica una relazione tra i valori della variabile indipendente e quelli della funzione non presenta difficoltà quando l'una e l'altra assumono valori discreti, come nell'esempio a cui si riferisce la formula (50), ma quando essi assumono invece valori continui è necessario stabilire la relazione tra classi di valori, indicando l'intervallo fra i valori estremi della classe.

Per la rappresentazione di tale intervallo sono in uso due metodi.

Secondo l'uso di essi, per indicare l'intervallo tra due quantità per esempio 54 e 56, si può scrivere 55 ($\pm$ 1) e per indicare l'intervallo tra 55 e 56 si può scrivere 55 ($+$ 1) o anche 56 ($-$ 1) o anche 54,5 ($\pm$ 0,5).

Questo metodo offre l'inconveniente che esso può dar luogo in certi casi a qualche confusione, in quanto esso viene usato con particolare significato per rappresentare l'errore probabile delle costanti statistiche, il valore della costante essendo indicato in tal caso dal numero fuori della parentesi, mentre fra parentesi vengono indicati i limiti negativo e positivo dell'errore probabile. Così l'espressione sopra indicata 55 (± 1) potrebbe prendersi come l'indicazione di una costante statistica il cui valore osservato è = 55 e di cui si ritiene che il valore vero cada con la probabilità = 1/2 tra 54 e 56.

Spesso però il pericolo di una tale confusione non si verifica ed allora il metodo è consigliabile in considerazione della sua brevità, mentre quando la variabile indipendente varia con uniformità, esso non presenta seri inconvenienti.

Il metodo è per esempio usato per enunciare importanti risultati del calcolo delle probabilità. Tale è quello che, se con p si indica la probabilità nota di un evento e con F la sua frequenza verificatasi in n osservazioni, è da attendersi, qualora p non sia troppo diverso da 1/2 ed n sia sufficientemente grande,

$$F \overset{0{,}683}{=} p\left(\pm \frac{p\,(1-p)}{n}\right)$$

e in generale

$$F \overset{z}{=} p\left(\pm t \frac{p\,(1-p)}{n}\right)$$

dove z è una funzione biunivoca di t.

Ancora più in generale, è dimostrato che, nel caso di un fenomeno Y a distribuzione gaussiana con scostamento quadratico medio σ_y, è da attendersi

$$Y \overset{z}{=} \bar{y}\,(\pm t\,\sigma_y)$$

vale a dire che una misura singola del fenomeno Y ha la probabilità z di allontanarsi dalla media aritmetica $\bar{y}$ per uno scostamento inferiore in valore assoluto a $t\,\sigma_y$. Apposite tabelle forniscono i valori di z in corrispondenza dei vari valori di t.

Il secondo metodo consiste nello scrivere, separati da una lineetta, i due limiti estremi dell'intervallo. Oltre al vantaggio di risparmiare la sia pur minima fatica di eseguire l'operazione

o le operazioni che nel primo metodo sono indicate tra parentesi, esso ha quello di permettere di precisare se il limite massimo o il limite minimo dell'intervallo od entrambi i limiti sono compresi od esclusi nell'intervallo stesso.

È questa una circostanza che può assumere grande importanza nei casi in cui i valori si accumulano sopra i valori limiti (in generale rappresentati da cifre rotonde), come avviene nel caso delle dichiarazioni di età dei censiti e ancor più nella fissazione dell'ammontare dei capitali delle società per azioni.

È necessario perciò che i due apici della lineetta che separa le cifre che indicano i due limiti dell'intervallo si differenzino a seconda che il limite sia compreso od escluso.

Si suole, in matematica, intendere che il limite sia escluso quando l'apice della lineetta non porta alcun segno ed invece compreso quando essa porta perpendicolarmente una piccola sbarra. Così la notazione

50 ⊢——— 55

indicherebbe che il 50 è compreso e il 55 escluso dall'intevallo.

La notazione sembra per verità poco rispondente, in quanto la sbarra perpendicolare suggerisce piuttosto che il limite resti escluso, e ciò può facilmente ingenerare equivoco in chi non ha dimestichezza con tale notazione.

Volendo modificare quanto meno si può la notazione in uso, io suggerirei di adottare, invece di una piccola sbarra, una piccola graffa e, d'altra parte, porre all'apice della lineetta un punto per indicare che il limite è escluso dall'intervallo.

Così la notazione

50)——• 55

significherebbe: da 50 compreso a 55 escluso, mentre la notazione

50 •——(55

significherebbe da 50 escluso a 55 compreso; ancora la notazione

50 •——• 55

significherebbe da 50 escluso a 55 escluso, ed infine la notazione

50)——(55

significherebbe da 50 compreso a 55 compreso.

Quando, come nei casi che qui si considerano, la variabile può assumere valori continui, i numeri con cui se ne indicano le intensità sono quasi sempre approssimati e l'approssimazione può ottenersi o per abbreviazione o per arrotondamento. Il numero non esprime così in realtà un valore singolo, ma una classe di valori più o meno ampia.

Ora, può essere importante precisare se si tratta di valore abbreviato o arrotondato; se, per esempio, 50 significa da 50 a 51 o invece da 49,5 a 50,5. Ciò è importante quando la variabile cresce o diminuisce rapidamente. Un rapporto fra il numero dei casi di un fenomeno relativo ad un numero arrotondato e il numero dei casi di un altro fenomeno relativi allo stesso numero abbreviato, può, in tali casi, portare a risultati profondamente erronei. Così è avvenuto nella determinazione dei coefficienti specifici di natalità legittima (rapporti dei nati legittimi alle donne coniugate di una data età) per singoli anni da 15 a 20 in un paese in cui le età delle coniugate date dal censimento e l'età delle madri data dalle statistiche delle nascite erano state esposte diversamente, l'una con un numero abbreviato e l'altra con un numero arrotondato.

Sarebbe sempre bene, in tali casi, che non si indicasse il numero, ma la classe che il numero rappresenta, o almeno si precisasse, in un'avvertenza, se il numero è stato abbreviato o arrotondato.

Per distinguere a prima vista la notazione di un intervallo da quella della sottrazione di due numeri, sarà bene porre la notazione dell'intervallo in una nicchia, nel modo seguente

$$\left|\,50 \rightarrow 51\,\right| \qquad \left|\,49{,}5 \rightarrow 50{,}5\,\right|$$

* * *

Nelle pagine precedenti, noi siamo ricorsi, per differenziare la variabile indipendente dalla funzione, al primo tra i due metodi prospettati.

Di fronte ad esso, il secondo presenta alcuni vantaggi.

Anzitutto esso permette di conservare un simbolo generico per la variabile (sia indipendente che dipendente) senza bisogno di ricorrere all'espediente, a cui noi siamo finora ricorsi, di scrivere, secondo la loro pronuncia, le lettere dell'alfabeto con cui

le variabili si esprimono (per esempio, « emme », « enne », « bi », ecc.) espediente che d'altra parte cade in difetto nel caso delle vocali.

In secondo luogo, esso si presta a una rappresentazione più semplice delle funzioni biunivoche. Per esempio, per esprimere che « emme » è funzione biunivoca di « esse », col secondo metodo si scriverà

$$m \leftrightharpoons s$$

anzichè ricorrere a due espressioni come si dovrebbe fare col primo metodo

$$M = s$$

$$S = m$$

In terzo luogo, il secondo metodo presenta un'analogia con metodi, che saranno esposti nelle pagine che seguono, adatti per indicare la corrispondenza tra quantità statistiche.

D'altra parte, il secondo metodo ha l'inconveniente di essere limitato alle uguaglianze. Per le disuguaglianze come la (46) e la (46.2) si dovrebbero adottare simboli corrispondenti, che pure indicassero la direzione con cui si passa dalla variabile alla funzione; per esempio, i seguenti, i quali sono già complicati, nel caso di funzioni univoche $\neq$ $>$ $<$ e i seguenti, ancora più complicati, nel caso di funzioni biunivoche $\neq$ $\gtrless$

* * *

Trovare formule semplici che esprimano matematicamente le relazioni che intercedono tra più, o anche tra due soli, fenomeni statistici non è cosa facile. Una formula naturalmente si può sempre trovare, ma, quando essa è complicata, a poco o a nulla serve nella pratica. D'altra parte le formule sono sempre approssimative : ne viene che le relazioni tra gli stessi fenomeni possono essere espresse, spesso con gradi di approssimazione diversi, ma talvolta con gradi di approssimazione presso a poco uguali, con formule differenti. Per esempio, la relazione fra statura e peso dei soldati può essere rappresentata con buona approssimazione con una parabola di secondo grado, secondo la proposta di Quetelet, del pari che con una retta [1]. Per alcune serie di dati

[1] Cfr. la nostra memoria *Delle relazioni fra le intensità cograduate di due caratteri* in « Atti del R. Istituto Veneto di Scienze, Lettere ed Arti », anno acc. 1916-17, vol. XXVI, parte II, pp. 1161-1162.

l'approssimazione è ugualmente buona con le due formule ; per altre essa risulta migliore con una retta. Ancora, come si è visto, queste formule non esprimono le relazioni delle intensità osservate di uno o più fenomeni con le intensità osservate di un altro fenomeno nei casi singoli, ma solo con la media delle sue intensità osservate in un sufficiente numero di casi e questa relazione d'altronde non è invertibile, cosicchè non si può dire di aver determinato la relazione tra le intensità dei fenomeni in questione.

Per tutte queste ragioni, l'attività degli statistici, più che a determinare formule generali che esprimano le relazioni che intercedono tra due o più fenomeni o caratteri, è diretta a descrivere empiricamente tali relazioni mediante tavole a semplice entrata, nel caso di relazioni tra due fenomeni o caratteri, o a entrata duplice, triplice, quadruplice, ecc., nel caso di relazioni fra tre, quattro, cinque o più fenomeni o caratteri. Effettivamente, solo per poche relazioni statistiche si sono determinate le formule empiriche, mentre tutte le pubblicazioni statistiche sono piene di tavole che presentano appunto le relazioni empiriche osservate tra due o più fenomeni o caratteri. Queste tavole esprimono le intensità concrete di un fenomeno in corrispondenza con le modalità concrete quantitative e qualitative di uno o più altri fenomeni. Ora è opportuno disporre di simboli adatti per indicare le intensità concrete dei fenomeni e per significare la loro corrispondenza.

Se a indica in generale la intensità in un caso concreto di un fenomeno A, con $a_1, a_2, a_3 \ldots a_n$ si sogliono indicare i valori che tale intensità assume in n casi concreti, indicati ed elencati secondo uno o un altro criterio, e, in generale, con a_i uno di tali valori o, come suol dirsi, il valore generico di A. Si avverta la differenza di significato tra a e a_1. Il primo indica una classe di intensità possibili, che possono essere in numero finito (per esempio nel caso del numero dei figli o del numero dei coniugi avuti) o infinito (per esempio, nel caso della altezza della statura, dell'età, ecc.) ; il secondo indica una intensità tra una classe di intensità verificatisi, sempre in numero finito.

I simboli della corrispondenza dovranno naturalmente distinguere tra corrispondenza biunivoca, univoca e plurivoca.

Useremo la notazione

$$a_i \overset{z}{\longleftarrow} b_i$$

per indicare che al valore concreto b_i del fenomeno B corrisponde con una frequenza z $(1 \geq z \geq 0)$ il valore concreto a_i del fenomeno A.

Viceversa

$$a_i \xrightarrow{z} b_i$$

indicherà che al valore concreto a_i corrisponde, con una frequenza r $(1 \geq r \geq 0)$, il valore concreto b_i.

I due simboli si possono riunire. La notazione

$$a \xleftarrow{z} b_i$$

indicherà che al valore concreto b_i corrisponde il valore concreto a_i con una frequenza z, mentre al valore concreto a_i corrisponde il valore concreto b_i con una frequenza r.

Quando z o r è uguale ad 1 si può omettere di indicarlo.

Quindi la relazione

$$a_i \longrightarrow b_i$$

indica che vi è corrispondenza univoca di b_i ad a_i, cioè che, quando si verifica a_i, si verifica sempre b_i, e analogamente la relazione

$$a_i \longleftarrow b_i$$

indica che vi è corrispondenza univoca di a_i a b_i, vale a dire che, tutte le volte che si verifica b_i, si verifica a_i, e infine la notazione

$$a_i \longleftrightarrow b_i$$

significa che vi è corrispondenza biunivoca tra a_i e b_i.

La notazione

$$a_i \xleftarrow{z} b_i$$

indicherà che vi è corrispondenza univoca di a_i a b_i, mentre al valore di a_i corrisponde solo con una frequenza r il valore di b_i, e la notazione

$$a_i \xleftarrow{z} b_i$$

indicherà che vi è corrispondenza univoca di b_i ad a_i, mentre al valore di b_i corrisponde solo con una frequenza z il valore di a_i.

Con l'aggiunta di due piccoli segni alla detta notazione si potrà altresì precisare se il valore corrispondente è il valore modale, cioè quello che si avvera con la massima frequenza, op-

pure se esso è superiore o inferiore a tale valore modale. Nel primo caso, porremo la lettera z o r tra due lineette convergenti; nel secondo caso, le porremo tra due lineette parallele inclinate da destra a sinistra o rispettivamente da sinistra a destra.

Quindi la notazione

$$a_i \; \angle \; b_i$$

significa che al valore concreto b_i corrisponde con una frequenza z il valore concreto a_i e che questo valore b_i è, tra tutti i valori corrispondenti ad a_i, il più frequente; la notazione

$$a_i \; \overline{} \; b_i$$

significa che al valore concreto a_i corrisponde con una frequenza r il valore concreto b_i e che questo è inferiore al valore di B, corrispondente ad a_i che ha la massima frequenza; la notazione

$$a_i \; \angle \; b_i$$

significa che al valore concreto b_i di B corrisponde con una frequenza z il valore concreto a_i di A, il quale a_i è superiore al valore di A corrispondente a b_i che presenta la massima frequenza, mentre al valore concreto a_i di A corrisponde con una frequenza r il valore b_i di B che, tra tutti i valori di B corrispondenti al valore concreto a_i, presenta la massima frequenza.

Indichiamo, per esempio, con s la statura espressa in cm. e con p il perimetro toracico pure espresso in cm. dei militari italiani di 20 anni delle classi 1859-63. Per la statura di cm. 161 ed il perimetro toracico di cm. 85 si può scrivere

$$85 \; (p) \; \angle \; 161 \; (s)$$

che ci dice che ad una statura di 161 cm. corrisponde nel 12% dei casi un perimetro toracico di 85 cm. e che tutti gli altri perimetri toracici corrispondenti a tale statura si presentano con minore frequenza, e viceversa che a un perimetro toracico di 85 cm. corrisponde nel 7% dei casi una statura di 161 cm., ma che la statura più frequente, corrispondente ai perimetri toracici di 85 cm., non è questa, ma una statura meno elevata [1].

---

[1] Per i dati, vedi *Antropometria militare*, Parte II, *Dati demografici e biologici*, Roma, 1903, Parte X, pagg. 92-93.

Naturalmente la corrispondenza si potrà stabilire non solo tra due valori singoli, ma anche tra un valore singolo e un valore medio. La notazione

$$a_i \longrightarrow \overline{b}_i$$

indicherà che al valore concreto a_i del fenomeno A, corrisponde il valore medio aritmetico $\overline{b}_i$ del fenomeno B e le notazioni

$$a_i \longrightarrow \overset{|}{b}_l$$

$$a_i \longrightarrow \hat{b}_k$$

indicheranno che al valore concreto a_i del fenomeno A corrisponde il valore mediano $\overset{|}{b}_l$ del fenomeno B e il valore modale $\hat{b}_k$ di detto fenomeno.

Due o più notazioni si possono riunire, scrivendo

$$a_i \longrightarrow \overline{b}_i \, , \, \overset{|}{b}_l \, \hat{b}_k$$

Per esempio, tra i suddetti militari italiani di 20 anni delle classi 1859-63, alla statura (s) di 161 cm. corrisponde un perimetro toracico (p) medio aritmetico di cm. 86,2, un perimetro toracico mediano di cm. 86,5, un perimetro toracico di massima frequenza di cm. 85 [1]. Tutto ciò si potrà scrivere molto semplicemente con la notazione

$$161 \ (s) \longrightarrow \overline{86,2} \ ; \ \overset{|}{86,5} \ ; \ \hat{85} \ (p)$$

Naturalmente queste relazioni non sono invertibili: così, mentre il perimetro toracico di 85 cm. costituisce il valore modale per la statura di 161 cm., la statura di 161 cm. non costituisce il valore modale per i perimetri torarici di 85 cm.; detto valore modale è, invece, di 157 cm., ciò che si indicherà scrivendo

$$85 \ (p) \longrightarrow \hat{157} \ (s)$$

---

[1] Vedi, per i dati, la citata *Antropometria militare*, Parte II Tav. X, a pagg. 92-93 e 94-97, Tav. XI a pag. 98.

oppure

$$\hat{157}\ (s) \longleftarrow 85\ (p)$$

Facciamo un altro esempio, relativo ai coscritti belgi della classe 1926 [1].

Ad un perimetro toracico di cm. 92 corrisponde una statura media aritmetica di cm. 170, ma ad una statura di cm. 170 corrisponde un perimetro toracico medio aritmetico di poco più di 85 cm. Potremo quindi scrivere

$$\overline{170}\ (s) \longleftarrow 92\ (p)$$

$$170\ (s) \longrightarrow \overline{85}(p)$$

Similmente a una statura di 174 cm. corrisponde un perimetro toracico medio aritmetico di 86 cm., ma a un perimetro toracico di 86 cm. corrisponde una statura media aritmetica di poco più 168 cm.

$$\overline{86}\ (p) \longleftarrow 174\ (s)$$

$$86\ (p) \longrightarrow \overline{168}\ (s)$$

Vediamo ancora un esempio, ricavato, questo, dalle statistiche economiche. Nel 1906, a un patrimonio (p) di 400.000 franchi corrispondeva a Zurigo un reddito medio aritmetico da lavoro (r) di 8.200 franchi, mentre un patrimonio medio aritmetico di 400.000 franchi corrispondeva a un reddito da lavoro di 45.600 franchi [2], così che potrà scriversi

$$400.000\ (p) \longrightarrow \overline{8.200}\ (s)$$

$$\overline{400.000}\ (p) \longleftarrow 45.600\ (r)$$

---

[1] Trattasi di 47.832 coscritti, studiati dai Dott. Govaerts e Sillevaerts. Vedi la memoria *Contribution à l'analyse de l'évolution de la robusticité chez les miliciens belges*, « Recueil des Mémoires couronnés et autres mémoires de l'Académie royale de Médecine de Belgique », tome XXIV, Bruxelles, 1930. Per l'elaborazione dei dati, vedi la nostra memoria, citata alla nota seguente.

[2] Vedi, per questi ed altri dati consimili, la nostra memoria *Delle relazioni tra le intensità cograduate di due caratteri*, in « Atti del R. Istituto Veneto di Scienze, Lettere ed Arti », Anno accademico 1916-17, vol. XXVI, Parte seconda, pag. 1155.

* * *

I simboli proposti possono servire ad esprimere la corrispondenza, non solo tra due valori concreti a_i e b_i o tra un valore concreto e un valore medio, ma anche tra un valore concreto a_i e un intervallo $|b'_i \rightarrow b''_i|$ o tra due intervalli $|a'_i \rightarrow a''_i|$ e $|b' \rightarrow b''_i|$.

Abbiamo anzi già avvertito che, quando si tratta di fenomeni o caratteri le cui misure variano con continuità — come è appunto il caso per la statura e il perimetro toracico — le statistiche in realtà non forniscono i valori precisi, ma i valori classificati per intervalli più o meno estesi. Le stature indicate come di 161 cm. in realtà sono stature comprese nell'intervallo 160,5 — 161,5 cm., oppure nell'intervallo 161 — 162 cm., a seconda del criterio adottato nella misura e nella classificazione; e analogamente i perimetri toracici indicati come di 85 cm. sono perimetri toracici compresi nell'intervallo 84,5 — 85,5 cm. oppure nell'intervallo 85 — 86 cm. Nel caso dei militari italiani, è il secondo criterio che si è adottato, almeno nella classificazione delle stature [1].

Invece di scrivere, come abbiamo fatto a pag. 32, in forma abbreviata

$$85\ (p) \overset{0,12}{\underset{0,07}{<}} 161\ (s)$$

si dovrebbe pertanto scrivere più esattamente

$$|85 \rightarrow 86|\ (p) \overset{0,12}{\underset{0,07}{<}} |161 \rightarrow 162|\ (s)$$

Analogamente per intervalli più ampi.

Per esempio, per i militari italiani sopra considerati, si trova che, alla statura di 161 cm. sono indicati come corrispondenti, nel 43,7% dei casi, perimetri del torace di 85, 86, 87, 88 cm. e, nel 69,6% dei casi, perimetri toracici da 82 cm. a 88 cm. (incluso), ciò che può esprimersi scrivendo

$$\left.\begin{array}{l} |85 \rightarrow (88| \overset{0437}{<} \\ |82 \rightarrow (88| \overset{0696}{<} \end{array}\right\} |161 \rightarrow 162|\ (s)$$

[1] Vedi *Antropometria militare*, Parte I: *Dati antropologici ed etnologici*, Roma, 1898. Nota a pag. 25.

L'espressione

ci dice che, a stature di cm. 160, 161, 162, corrispondono, nel 33,9% dei casi, perimetri toracici di cm.84, 85, 86 e che, a perimetri toracici di 84, 85, 86 cm., corrispondono, nel 23,8% dei casi, stature di 160, 161, 162 cm.

* * *

La corrispondenza tra le quantità di due gruppi può essere stabilita in base a criteri vari: in generale viene stabilita mediante criteri sostanziali: per esempio, in base al criterio di coesistenza di due caratteri in uno stesso organismo o di due persone in uno stesso gruppo, come nel caso della corrispondenza tra statura e perimetro toracico dello stesso individuo o tra statura della moglie e statura del marito; oppure in base al criterio di discendenza, come nel caso della corrispondenza tra la statura del padre o del nonno e la statura del figlio o rispettivamente del nipote. La corrispondenza può essere però stabilita anche con criterio formale: per esempio, tra le quantità che occupano lo stesso posto nelle graduatorie, entrambe crescenti o entrambe decrescenti, di due gruppi di quantità ugualmente numerose. In questo secondo caso, si parla di *cograduazione*, mentre, quando si parla di *corrispondenza*, si intende, per antonomasia, riferirsi alla corrispondenza stabilita in base a criteri sostanziali.

È bene introdurre simboli appropriati anche nel caso della cograduazione.

Cograduate si dicono due quantità che occupano lo stesso posto in due graduatorie, entrambe crescenti o entrambe decrescenti. Per indicare che un dato valore a_i di un gruppo A ed un dato valore b_i di un altro gruppo B sono cograduati può usarsi la notazione

$$a_i \longleftrightarrow b_i$$

Contrograduate si dicono invece due quantità che occupano lo stesso posto nelle graduatorie contrarie — una, cioè, crescente e l'altra decrescente — delle quantità di due gruppi ugualmente

numerosi. E, per indicare che due quantità a_i e β_i sono contrograduate, potrà usarsi la notazione

$$a_i \longleftrightarrow \beta_i$$

La quantità cograduata ad un'altra non coincide sempre con la quantità media corrispondente; essa ne può restare superiore o inferiore. Per esempio, alla statura (s) di 155 cm. risultava cograduato, per i soldati italiani, un peso (p) di 47 Kg., ma corrispondeva un peso medio aritmetico di Kg. 55,1; a una statura di 185 cm. risultava cograduato un peso di Kg. 85,5 e corrispondeva un peso medio aritmetico di Kg. 71 [1], ciò che si può esprimere scrivendo

$$155\ (s)\begin{cases}\longleftrightarrow 47\ (p)\\ \longrightarrow \overline{55{,}1}\ (p)\end{cases}$$

$$185\ (s)\begin{cases}\longleftrightarrow 85{,}5\ (p)\\ \longrightarrow \overline{71{,}0}\ (p)\end{cases}$$

Per i coscritti belgi, abbiamo visto che a una statura di 174 cm. corrispondeva un perimetro medio aritmetico di 86 cm.; il perimetro cograduato è invece di cm. 89,2. A un perimetro di 86 cm. corrispondeva una statura media aritmetica di cm. 168; il perimetro toracico cograduato a una statura di 168 cm. è invece di cm. 84,6.

Potremmo scrivere

$$174\ (s) \longleftrightarrow 89{,}2\ (p)$$

$$174\ (s) \longrightarrow \overline{86}\ (p)$$

$$\overline{168}\ (s) \longleftarrow 86\ (p)$$

$$168\ (s) \longleftrightarrow 84{,}6\ (p)$$

Analogamente si è visto che, per gli stessi soldati, a un perimetro toracico di 92 cm. corrispondeva una statura media aritmetica di 170 cm. mentre a una statura di 170 cm. corrispon-

---

[1] Vedi, per questi ed altri dati, la citata memoria *Delle relazioni tra le intensità cograduate di due caratteri*, nota [1] a pag. 1163.

deva un perimetro toracico medio aritmetico di 85 cm. Ora, il il perimetro toracico cograduato alla statura di 170 cm. è di 86 cm., intermedio tra i due valori precedenti, così che si scriverà :

$$170\ (s) \longrightarrow \overline{85}\ (p)$$
$$170\ (s) \longleftrightarrow 86\ (p)$$
$$\overline{170}\ (s) \longleftarrow 92\ (p)$$

Abbiamo visto che a Zurigo, nel 1906, a un patrimonio di 400.000 franchi corrispondeva un reddito medio aritmetico da lavoro di 8.200 franchi, mentre un patrimonio medio aritmetico di 400.000 franchi corrispondeva ad un reddito da lavoro di 45.600 franchi. Molto diverso sia da 8.200 che da 45.600, e precisamente uguale a 10.700, risultava il reddito da lavoro cograduato ad un patrimonio di 400.000 franchi (1), ciò che si potrà esprimere scrivendo

$$400.000\ (p) \longrightarrow \overline{8.200}\ (r)$$
$$400.000\ (p) \longleftrightarrow 10.700\ (r)$$
$$\overline{400.000}\ (p) \longleftarrow 45.600\ (r)$$

In questo caso, come nel precedente, il reddito cograduato risulta intermedio, ma non è detto che debba essere sempre così.

Per il patrimonio e il reddito globale (in corone) in Danimarca, nel 1908, risulta (2).

$$3.000\ (p) \longrightarrow \overline{930}\ (r)$$
$$3.000\ (p) \longleftrightarrow 1.200\ (r)$$
$$\overline{3.000}\ (p) \longleftarrow 1.100\ (r)$$

---

(1) Vedi Op. cit. luogo citato.
(2) Vedi Op. cit. pag. 1156.

Résumé

L'auteur rappelle, au préalable, certains problèmes de statistique méthodologique et appliquée, en traitant lesquels des malentendus sont nés, qu'on aurait pu éviter en ayant recours à l'emploi de symboles appropriés. L'auteur propose, en conséquence, de nouveaux symboles, qui pourraient être utilement employés dans les recherches statistiques ; il ne soutient pas qu'il faudrait les utiliser toujours, mais seulement lorsque le fait de n'y point recourir pourrait donner lieu à des malentendus.

Il rappelle, tout d'abord, la distinction connue des « égalités » en « identités » et « équations », et aussi bien parmi les unes que parmi les autres, il distingue les « déterminées », qui valent seulement pour une ou pour certaines valeurs données des inconnues, et les « indéterminées », dans lesquelles chacune des inconnues peut prendre des valeurs diverses. Parmi ces dernières, il fait encore une distinction entre « illimitées », dans lesquelles chacun des termes peut prendre n'importe quelle valeur, « presque illimitées », dans lesquelles ces inconnues peuvent prendre n'importe quelle valeur, excepté quelques valeurs spéciales données, toujours en nombre fini, et « limitées », dans lesquelles les inconnues peuvent prendre seulement des valeurs qui rentrent dans des catégories particulières (par exemple, positives ou négatives). Pour chacune de ces catégories et sous-catégories d'équations ou d'identités, de même que pour les « inégalités » correspondantes, l'auteur propose des symboles appropriés.

Passant ensuite en particulier aux équations indéterminées, il introduit la notion d'« ordre de l'indétermination » et distingue les relations biunivoques, univoques, plurivoques et biplurivoques, mettant en évidence le divers degré d'invertibilité qui revient à chacune d'elles.

De l'examen de cette question, il découle l'opportunité d'adopter des symboles dûment choisis, que l'auteur introduit, en vue de distinguer :

a) la variable de la fonction ; *b*) le rapport entre variable et fonction du rapport entre variable et valeur moyenne de la fonction ; *c*) les plus importantes valeurs moyennes de la fonction (moyenne arithmétique, médiane, mode, antimode) et, à leur opposé, les valeurs individuelles de la fonction ; *d*) la fréquence,

avec laquelle s'avère une relation, de la fréquence de la relation inverse ; *e*) les classes de valeurs d'une variable continue des valeurs d'une variable discrète ; *f*) les relations invertibles des relations sub-invertibles. L'auteur présente des exemples susceptibles de mettre en évidence la portée de ces distinctions.

Il passe, successivement, à mettre au jour l'importance qu'assument dans la pratique, la description empirique des phénomènes statistiques ; et, à cet effet, les relations de correspondance entre leurs valuers. Des notations appropriées sont proposées, en vue de représenter ces relations et de montrer le sens et la fréquence suivant lesquels s'établit la correspondance.

L'auteur distingue enfin, les relations de correspondance, celles de cograduation et de contre-graduation, proposant aussi pour celles-ci des symboles appropriés.

S. R. DAS

A mathematical analysis of the phenomena of human twins and higher plural births. Part I: Twins.

The purpose of the present paper is to give a mathematical analysis of the phenomena of multiple births, based on well-established biological principles. Four biological quantities a, μ, Φ and D which are explained below, have been introduced and the whole structure of the analysis has been built up with them. The chief biological principles and facts which form the background of the reasonings underlying the mathematical analysis are discussed at some length in the sections 1-10.

(1) Two types of twins

There are two different types of twins, "identical" or monozygotic and "non-identical" or dizygotic. The former comes into existence by a twinning division of a single zygote, formed by the fertilization of a single ovum by a single mature sperm and the latter type of twins, are formed as a result of the fertilisation of two independently liberated ova by two separate mature sperms; they may, therefore, have like or unlike sexes and are generally not more similar than two sibs born of the same parents. The members of a monozygotic twin pair must, of necessity, be of the same sex and are very similar in appearance. A pair of monozygotic twins are of identical genotype and very similar phenotypes, a dizygotic twins pair have different genotypes as well as phenotypes.

This is however too simple a picture of the phenomena that actually happen. The mutual interaction of the genes and the

influence of the environments in the determination of the phenotypes render twin diagnosis a much complicated problem.

Price (1950) has drawn more serious attention to the prenatal and natal difference-producing factors in the case of monozygotic twins and has discussed three kinds of biases affecting the conclusions of twin studies.

(2) Genesis of monozygotic twins

Monozygotic twinning is supposed to take place as a result of an inherent tendency in certain zygotes to undergo a twinning division (Dahlberg '26) at an early stage of their life. This tendency may not, in some cases, be satisfied after one such division i.e. after the formation of a pair of twins, but it may persist even further and the one, or the other, or both of the members of the twin pair may again undergo similar scissions producing monozygotic triplets, quadruplets, etc. Thus, in order that a monozygotic twin pair should result, the twinning division must take place once and once only. In order that a monozygotic triplet should result, the splitting must take place twice : the first division of the single zygote produces an " identical " pair, one and only one member of which should then similarly split up into two, but once. Similar pictures of the processes for the formation of higher monozygotic multiplicity can be visualised.

(3) Genesis of dizygotic twins

The genesis of dizygotic twins is altogether different from that of the monozygotic twins. Two ova liberated from one, or both of the ovaries within a small interval of time and fertilised by two different sperms lead to the formation of two zygotes which develop and grow inside two separate choria with separate plancentae and thus form a pair of dizygotic twins. Dizygotic twins are therefore expected to be no more similar than two sibs. They may be either like-sexed or unlike-sexed. The proportions of male and female homo-sexual and the hetetro-sexual pairs among a population of dizygotic twins follow certain statistical rules depending upon the sex-ratio of the population and the probabilities of the two types of twinning in all births.

(4) The two types of twins are therefore the manifestations of two basically distinct phenomena (Dahlberg, '26 ; Greulich, '35) and are not merely the different expressions of one and the same twinning tendency. This well-founded opinion are definitely against the hypotheses introduced by Fisher ('19) and Jenkins ('29).

(5) A third type of twins having identical heredity from the maternal side and different heredity from the side of the father has sometimes been hypothesised from mere theoretical speculations (Danforth, '16 and Fisher, '19). As this idea has never found favour with the human biologists for want of proper evidence, it has now been practically discarded. The present paper therefore has not taken any account of this hypothetical twins in the development of its theories.

(6) Mothers' age and births statistics

An interesting relation between the age of the mother and the frequency of twin confinements relative to that of all confinements has been reported by a number of authors (Dahlberg, '26 ; Jenkins '27 & '29 ; Greulich, '30 ; Komai and Fukuoka, '36 ; Enders and Stern, '48). Sarkar ('44) collected some Indian data and proved a similar relationship between the frequency of unlike-sexed twins and the age of the mothers in the case of 84 twins. Enders and Stern ('48) have concluded (i) that the chance of a mother, both whites and Negroes, having dizygotic twins confinements increases steadily with the age of the mother from the age group (15-19), up to the (35-39) group and then abruptly the curve slopes downwards in the (40-44) class ; (ii) that the chance for occurrence of monozygotic twin confinements shows a slight but steady tendency to increase with the age of the mother, both Whites and Negroes, throughout the whole reproductive range of age ; (iii) that the age-rate of increase of the chance for monozygotic twin confinements in the case of the Whites is significantly different from that of the Negroes, and (iv) that the fact (ii) is indicative of factors favouring either the occurrence of twinning in zygotes during their early development, or the viability of the twin embryos, or the combination of the two.

Dahlberg, Jenkins and Greulich's European data and Komai and Fukuoka's Japanese data agree quite closely in this respect with the recent observations of Enders and Stern based upon the more accurate American data. Sarkar ('44) did not give the frequency of all confinements at the various age-groups of the mothers and hence the actual trend of the increase of chance for twin confinements cannot be found out from those data.

In the light of what have been stated here in this connection, it therefore seems more rational, or rather absolutely essential that a mass of twin data should, first of all, be resolved, in accordance with the age-groups of the mothers into a number of divisions. Each division corresponding to a certain age-group of mothers should then be analysed further, say for example, by the application of the differential formula. Enders and Stern ('48) and Jenkins ('29) have exactly done so in comparing the twin and triplets birth statistics. We must remember that races may differ either in the ratio of the frequencies of the monozygotic to the dizygotic twin confinements, or in the frequency distribution of the twin confinements over the various age-groups of the mothers, or in both. It is also very important to express each frequency of twin confinement as percentage of the frequency of all confinements belonging to the same age-group of mothers.

For the purpose of the present paper we, therefore, assume that the probability of occurrence of a monozygotic or a dizygotic twin confinement among all confinements is dependent upon mother's age, or put in a more precise mathematical language — is a function of mother's age and that the function is different for the different races.

(7) Heredity and twin births

Heredity in the frequency of occurrence of twins has been the subject matter of study of a number of investigators (Weinberg, '01 ; Dahlberg, '26 ; Greulich, '35 ; Danforth, '16 ; Galton, 1875). We shall, in the present paper, take no positive account of this factor in developing the theory, as it can be supposed without any appreciable error that the parents, or families possessing a genetic disposition in favour of producing twins are more or less uniformly or randomly distributed in the popu-

lation and therefore when fairly large random samples of total confinements or all births are considered, not much variation is expected to appear in the results of statistical analysis and the probabilities of twin or higher plural confinements appear to be quite a steady characteristic of the population, or the race.

(8) Prenatal mortality

We cannot expect to get a satisfactory statistical record of the cases of abortions. Therefore when we speak of the total births, or all products of pregnancies, certain amount of inaccuracy is bound to be present. The inaccuracy is more serious than apprehended generally due to the fact that there is no record yet of the loss of fertilised ova that fail to survive and lose their lives in the utero within the first few weeks. At best we have at present to include the live-births, the still-born and the recorded cases of abortions in counting the total births. This is a weak point from which any theory including the present one based on these data shall suffer. The difficulty is further accentuated due to lack of an internationally agreed procedure in the records of the products of pregnancies. In countries like India where birth registration is not compulsory the data are poorer still in respect of their scientific utility (Sarkar, '44). The conditions prevailing in the advanced countries even (Strandskov, '45 ; Dahlberg, '26) are not entirely satisfactory. Thus a huge amount of data of immense importance to human biology are lowered in worth the world over for lack of scientific precision and uniformity in the method of registration.

(9) The sex-ratio – primary and secondary

The prenatal mortality rate for the male exceeds that of the females. It is for this reason that the secondary sex ratio (i.e. sex-ratio at birth) is less biased in favour of males than the primary sex-ratio which represents the number of (XY) zygotes to 100 of (XX) zygotes – immediately after the fertilisation.

It is again imperative here on our part to arrange the total confinements data according to the ages of the mothers.

(10) Multiple births sex-ratio

The sex-ratios among the various multiple birth confinements have been found to be less biased in favour of the male sex (Strandskov and Siemans, '46 *b* ; Wentworth, '14 ; Hamlett, '35 ; Dahlberg, '26 ; Sarkar, '44 ; Komai and Fukuoka, '36) and the higher the order of multiplicity, the lesser is the numerical predominance of the male sex. Several reasons have been suggested to explain the phenomenon, but none appears as conclusive (Strandskov and Siemens, '46). Strandskov and Siemens ('46) have considered two reasons for this phenomenon :

(1) " The male embroys under competition or under population pressure, tend to be eliminated to an even greater extent than do female embryos ".

(2) " The female determined zygote has a greater tendency to twin, or triple, or quadruple, than does the male determined zygote ".

If the first explanation be the correct one, " then the greater tendency for the elimination of the male embryo among plural pregnancies must occur early in pregency and not during the later stages, because even among still-births the percentage of males decreases with each increase in number per pregnancy. Most of the recorded still-births, it must be recalled, represent advanced stages in development ".

If the second explanation be the correct one, then the percentage of males will be lower only among the monozygotic plural pregnancies, but it will be reflected in the results of the total plural pregnancies. The authors have, however, favoured the first possibility without, of course, much conviction and stating any reasons against the second one. Thus they write " Although the second suggested hypothesis is a possibility, and the facts agree with it throughout, we are inclined to favour the first one. We must admit, however, that we cannot advance any deciding arguments in its favour ".

The present writer considered such a judgment in favour of the first and its acceptance quite arbitrary and thought a strict mathematical test would be the most rational way to our conclusion. In the absence of any quantitative study of the effects of the two alternative causes upon the expected plural

birth data and before testing how these expected values fit the observed ones, no really conclusive opinion can be given. It may be, that both these causes work jointly in deciding the observed results. In the present paper, we have introduced a differential twinning tendencies between the (XX) and (XY) zygotes in the probability factors Φ and μ which explain the twin data exactly or, rather Φ and μ have been so selected as could explain the sex-ratio among the twin confinements. A real test of the theory will, therefore, lie in the examination how the expected triplet or quadruplet sex-ratios fit the respective observed sex-ratios. It has been found on analysis of the American data reported by Strandskov and Siemens ('46), and Strandskov and Edelen ('46) that the same values of Φ and μ can partly, but not wholly account for the lower male percentage among the triplets. One defect in this analysis — a serious one too — is that the data were not classified according to the ages of the mothers.

We reserve these considerations and the consideration of the first hypothesis or their joint action for later communications.

(11) Hellin's law

Since Hellin (1895) formulated his empirical rule, almost all the reports on twin and higher multiple births statistics have referred to it. Hellin's law may be put as follows : The frequency of twin confinements bears to that of total confinements a ratio which is equal to the ratio borne by the frequency of the triplet confinements to that of the twin confinements. No report after Hellin has perhaps agreed satisfactorily with this rule. We may write the ratio as $1 : R$, where R is a racial constant. Strandskov ('45 *a*) has clearly shown the racial difference of R between the American " whites " and the Negroes. He has further shown with the help of χ^2 — test of the goodness of fit that the frequencies of the total, the twin, the triplet and the quadruplet confinements in the American " whites " and ' coloured " population (1922-36) do not agree with the expectations based on Hellin's law. Similar discordance with the law has been noted by almost all the analysts of plural births statistics (Komai and Fukuoka, '36 for Japan ; Das, '34 for India and Sarkar,

'45 for India and Ceylon). Gini ('23) has published data for more than 300 millions births, showing that Hellin's law, inexact for triplets, does not stand for quintuplets. The analysis presented in this paper proves that the Hellin's law has no sound basis. It is for this reason that exceptions to the rule has been the rule. Zeleny ('21) formulated a rule similar to Hellin's. Peller ('46) has attempted an improved formulation with certain modifications to Hellin's law. Jenkins ('27) indicated at first a theoretical support to Hellin's law and later on (1929) introduced an new hypothesis of 2-egg-twins formation to explain the triplet sex-homogeneity index. The hypothesis introduced by Jenkins ('29) assumes that dizygotic twin formation may take place by the fission of an egg before the reduction division into two which then undergo the reduction and fertilisation and form the twins. This idea like the one suggested by Fisher ('19) has no basis of biological observations. Jenkins and Gwin ('40) attempted a rigorous treatment of the live-births statistics with the classification of the data according as the age-groups of the mothers. The rule, so derived evidently deviates considerably from the Hellin's law. Gini ('24) has shown what hypotheses are implied in Hellin's law and how they should be modified in order that the observed data become in agreement with the theoretical ones.

(12) The differential formula

The differential formula which enables us to calculate the relative frequencies of the monozygotic and the dizygotic twin pairs out of the given frequencies of the ♂♂, ♀♀ and ♂♀ pairs, has proved to be the most useful tool to all students of twin births statistics. We must count all twin confinements, both live, one live, and both still-born cases for this purpose. The data should also be arranged according to the mothers' age. The differential formula is really a statistical device to adjust the proportions of the two types of twin frequencies in such a way as to exactly fit the data on the frequencies of the like- and the unlike-sexed twin pairs. It offers the best that can be done with the informations available. Some refinement is introduced by the consideration of the sex-ratio. Some use the sex-ratio of all confinements taken together and some again prefer to use the

sex-ratio among the twins alone. Strandskov and Edelen ('46) have critically compared both the procedures.

In the present paper we have shown in the following sections that the primary sex-ratio of the zygotes (XX) and (XY) before any loss due to death, or any plurality due to twinning processes take place after the fertilisation of the single ova-appears in the formula as a matter of course. Price (1950) has pointed out that the differential formula which popularly goes by the name of Weinberg ('01), was also known to Galton ('75) and Louis-Adolphe Bertillon (1874) was the first to devise it.

(13) The mathematical analysis

Let there be N single Zygotes formed in N pregnant mothers belonging to one of the age-groups into which the mothers are subdivided (see sec. 6). There will be two types of zygotes in respect of the sex-determining chromosomal constitutions: the (XX) and the (XY) types. The former develop into the female and the latter into the male embryos (Winnivarter, '12; Painter, '23, '24; Oguma and Kihara, '23; Evans and Swezy, '29; Crew, '46).

The number of the (XY) zygotes as a rule exceeds that of the (XX) zygotes and their ratio immediately after their formation is called the primary sex-ratio.

Let the primary sex-ratio $= \frac{a}{b}$, and $a + b = 1$.

The quantities a and b cannot be determined directly (section 9). Among the N zygotes, we must have,

the number of (XY) type $= Na$

and the number of (XX) type $= Nb$.

No definite answer to the question of numerical disparity between the two sex-determining genetic constitutions is yet available, although a few feasible explanations have been suggested. (Strandskov, '45 *b*; Strandskov and Bisaccia, '49; Schultz, '18; Parkes, '26).

As defined above, the quantities a and b can be alternatively put as follows:

$a =$ the probability of a zygote being of the $x\,y$ type and

$b =$ the probability of a zygote being of $x\,x$ type.

These two probabilities are mutually exclusive, but complementary.

Therefore, $a + b = 1$.

Hence, only one of a and b is really independent. We take a to be the independent variable, and $b = 1 - a$.

The quantity a is an important biological quantity and exhibits racial difference (Strandskov and Siemens, '46; Strandskov, '45).

(14) A few of the N single zygotes lose their singularity in two independent ways: namely, by the monozygotic and the dizygotic twin forming processes as explained in sections 1-3. These same processes lead by repetitions and combinations to plural pregnancies of different types from twins to higher orders of multiparities. We also take N as sufficiently large, so that the sample of N zygotes may represent the population without appreciable sampling variation.

Let,

μ = the probability of monozygotic scission of a XY-zygote;

Φ = the probability of monozygotic scission of a XX-zygote and

D = the probability of generation of an additional independent zygote giving rise to dizygotic twinning.

It follows therefore, that

$1 - \mu$ = the probability that a XY-zygote does not undergo any twinning division;

$1 - \Phi$ = the probability that a XX-zygote does not undergo any twinning division

and $1 - D$ = the probability that an additional and independent zygote giving rise to dizygotic twinning is not produced.

We can now evaluate the numbers of single XY- and XX-zygotes that are left after the short period during which the two twinning processes can work and do actually produce a few cases of plural confinements. This period of course may not necessarily be so short as might be thought (Stern, '49 and Price, '50, also Newman, '16 and Morril, '19).

Neglecting any loss of single zygotes by death during this short period, we may write:

$Na\ (1-\mu)\ (1-D)$ = the number of single XY-zygotes and

$Nb\ (1-\Phi)\ (1-D)$ = the number of single XX-zygotes.

If the differential mortality rates of the two types of zygotes at such an early stage of their lives can be ascertained in any way, and if we assume, then, that

δ = the probability of death of a XY-zygote during the period, in question, and

Δ = the probability of death of a XX-Zygote during the same period,

it follows further, that

$Na\ (1-\mu)\ (1-D)\ (1-\delta)$ = the number of single surviving XY-Zygotes, $= m_{11}$, say, and

$Nb\ (1-\Phi)\ (1-D)\ (1-\Delta)$ = the number of single surviving XX-Zygotes, $= f_{11}$, say.

Therefore, the sex-ratio of the single confinements after the completion of all plurality is given by

$$\left(\frac{m}{f}\right)_{11} = \frac{m_{11}}{f_{11}} = \frac{a}{b} \cdot \frac{(1-\mu)\ (1-\delta)}{(1-\Phi)\ (1-\Delta)},$$

or, neglecting the mortality factors,

$$\left(\frac{m}{f}\right)_{11} = \frac{a}{b} \cdot \frac{(1-\mu)}{(1-\Phi)}.$$

The suffix 11 of $\left(\frac{m}{f}\right)$ indicates 1-egg and 1-embryo confinements. We assume that counting all products of single confinements m_{11} and f_{11} can be determined from the birth statistics of the population (Strandskov, '45 *b*); but we have no easy means to determination of the sex of fetuses below 2 months of age and hence an amount of unavoidable error remains there to vitiate our estimates (Stern, '49) of m_{11} and f_{11}, even when it is taken for granted that all products of confinements are duly registered.

There is, of course, a method by which δ and Δ can be determined. The differential mortalities of the two sexes are known with some accuracy for the older fetuses of 4-10 months of age. From these data the mortalities at lower ages may be determined to the best estimate by extra-polation even though the mortalities seem to follow no regular mathematical law. We shall not in the present paper introduce these complexities (See sections 8 and 9).

The total number of single births is given by

$$N_{11} = N\,[a(1-\mu)+b\,(1-\Phi)]\,(1-D) \qquad [\because a+b=1]$$
$$= N\,(1-D)\,(1-a\,\mu-b\,\Phi)$$

(15) Monozygotic twins

As explained in section 2, if a single zygote undergoes twinning division once and once only and if no additional zygote is formed by the dizygotic twinning process (section, 3), a pair of monozygotic twins shall result. In order to find out the number of such cases among the XY- and XX-zygotes in our sample of N zygotes, the above conditions must be properly taken into account : Thus ;

$N\,a\,\mu\,(1-\mu)^2\,(1-D) =$ the number of ♂♂ pairs and

$N\,b\,\Phi\,(1-\Phi)^2\,(1-D) =$ the number of ♀♀ pairs.

Here, Φ and μ provide that a scission takes place and two twin zygotes are formed, none of which should undergo similar division and that condition is provided by the two factors $(1-\mu)$ $(1-\mu)$ or $(1-\Phi)$ $(1-\Phi)$ as the case may be. The factor $(1-D)$ imposes the condition that no dizygosity can occur.

The total number of monozygotic twins pair is given by

$$N_{12} = N\,(1-D)\,[a\,\mu\,(1-\mu)^2 + b\,\Phi\,(1-\Phi]^2$$

And, the sex-ratio among the monozygotic twin confinements is given by

$$\left(\frac{m}{f}\right)_{12} = \frac{a}{b}\cdot\frac{\mu\,(1-\mu)^2}{\Phi\,(1-\Phi)^2} = \left(\frac{m}{f}\right)_{11}\cdot\frac{\mu\,(1-\mu)}{\Phi\,(1-\Phi)}\,.$$

The suffix 12 indicates 1-egg and 2-embryos.

If we assume $\Phi > \mu$, then it can easily be proved that

$$\frac{\mu(1-\mu)}{\Phi(1-\Phi)} < 1$$

Hence,

$$\left(\frac{m}{f}\right)_{12} < \left(\frac{m}{f}\right)_{11}$$

i.e. the sex-ratio among 1-egg twins is less biased in favour of males than among the single confinements.

(16) Dizygotic twins

In section 14 we have explained the symbol D. But D may be written as

$$D = D\,a + D\,b, \qquad [\because a + b]$$

Where, $D\,a$ = the probability of appearance of a dizygotic male co-twin and

$D\,b$ = the probability of appearance of a dizygotic female co-twin.

Hence,

$N\,a\,D\,a\,(1-\mu)^2\,(1-D)$ = the number of ♂♂ 2-egg twin pairs, in which an initial male zygote finds a male co-twin ;

$N\,a\,D\,b\,(1-\Phi)\,(1-\mu)\,(1-D)$ = the number of ♂♀ 2-egg twin pairs in which an initial male zygote finds a female co-twin ;

$N\,b\,D\,a\,(1-\Phi)\,(1-\mu)\,(1-D)$ = the number of 2-egg ♀♂ twin pairs in which an initial female zygote finds a male co-twin, and

$N\,b\,D\,b\,(1-\Phi)^2\,(1-D)$ = the number of 2-egg ♀♀ twin pairs in which an initial female zygote finds a female co-twin.

These four permutations exhaust all the possible modes of dizygotic twin formation. The factors $D\,a$ or $D\,b$ provides the probability of occurrence of the independent additional zygote ; $(1-\mu)$ or $(1-\Phi)$ considers that no monozygotic scission takes place with either of the co-twins and the factor $(1-D)$ considers that no zygosity higher than 2 shall occur.

The total number of 2-egg pairs is given by

$$N_{22} = N\,D\,(1-D)\,(1-a\,\mu-b\,\Phi)^2.$$

Also,

$$\left(\frac{m}{f}\right)_{22} = \frac{a^2(1-\mu)^2 + a\,b\,(1-\mu)(1-\Phi)}{b^2(1-\Phi)^2 + a\,b\,(1-\mu)(1-\Phi)}$$

$$= \frac{a(1-\mu)}{b(1-\Phi)},$$

$$= \left(\frac{m}{f}\right)_{11}, \text{ from section 14.}$$

(17) Sex-ratio among total twin confinements

$$\left(\frac{m}{f}\right)_{12+22} = \frac{m_{12} + m_{22}}{f_{12} + f_{22}}$$

$$= \left(\frac{m}{f}\right)_{11} \cdot \frac{\mu(1-\mu) + D(1 - a\,\mu - b\,\Phi)}{\Phi(1-\Phi) + D(1 - a\,\mu - b\,\Phi)}.$$

Since, according to section 15,

$$\frac{\mu(1-\mu)}{\Phi(1-\Phi)} < 1$$

then,

$$\frac{\mu(1-\mu) + D(1 - a\,\mu - b\,\Phi)}{\Phi(1-\Phi) + D(1 - a\,\mu - b\,\Phi)} < 1$$

i.e. sex-ratio among all twin confinements is proved to be less biased in favour of males than among the single confinements and this is what is observed in the birth statistics (sec. 10).

If

$$\mu = \Phi, \text{ i.e. if } XY\text{- and } XX\text{-zygotes}$$

have the same probability for monozygotic twinning division, then it can be easily seen from sections 14-17 that

$$\frac{a}{b} = \left(\frac{m}{f}\right)_{11} = \left(\frac{m}{f}\right)_{12} = \left(\frac{m}{f}\right)_{22} = \left(\frac{m}{f}\right)_{12+22}$$

But when,

$$\Phi > \mu$$

we get,

$$\frac{a}{b} < \left(\frac{m}{f}\right)_{11} = \left(\frac{m}{f}\right)_{22} > \left(\frac{m}{f}\right)_{12+22} > \left(\frac{m}{f}\right)_{12}$$

(18) Relative frequencies of the sex combinations ♂♂, ♂♀ and ♀♀ among the dizygotic twins are given by the ratio,

$$a^2 (1 - \mu)^2 : 2 a b (1 - \mu) (1 - \Phi) : b^2 (1 - \Phi)^2$$

which is same as

$$(m_{11})^2 : 2 (m f)_{11} : (f_{11})^2 .$$

Among all twin confinements the same relative frequencies are given by the ratio,

$$a (1 - \mu)^2 (\mu + a D) : 2 a b D (1 - \mu) (1 - \Phi) : b (1 - \Phi)^2 (\Phi + b D) .$$

In twins birth statistics, this ratio is approximately put as 1 : 1 : 1 for Europeans, Americans and Indians, and 2 : 1 : 2 for Japanese (Komai and Fukuoka, '36); although in reality such simple ratios are never obtained.

The above general expression for the ratio may be much simplified, by the sacrifice of precision, thus if we put $\mu = \Phi$, and $a = b$ the above ratio stands as

$$(\Phi + b D) : 2 b D : (\Phi + b D)$$

If we further assume that

$$\mu = \Phi = a D = b D = \frac{1}{2} D ,$$

then the ratio becomes 1 : 1 : 1, the most popular ratio.

It is, therefore, clear that the conditions necessary for obtaining this simple ratio can never be satisfied. Hence, the ideal conditions necessary for the simple ratio are (1) that the primary sex-ratio must be 1 : 1 and (2) that the monozygotic twinning

probability must be half as much as the probability of occurrence of dizygotic twinning.

Similarly, the Japanese ratio 2 : 1 : 2 may be realised when

$$\Phi = \mu = \frac{3}{2} D \text{ and } a = b.$$

Hence, the necessary ideal conditions are (1) that the primary sex-ratio must be 1 : 1 and (2) that the monozygotic twining probability is half as much again as the probability of occurrence of dizygotic twinning.

(19) Frequencies of the monozygotic and the dizygotic twins

Let r = the frequency of the 2-male twins;

s = the frequency of the (1-male + 1-female) twins

t = the frequency of the 2-female twins,

where the expressions for r, s and t have been deduced in sections 15 and 16. The twin statistics offer the values of r, s and t which can be used in determining the relative proportions of the 1-egg and the 2-egg twins. From the expressions, deduced in 15 and 16, it is clearly found that r is divided into 1-egg and 2-egg twins in the ratio given by

$$r_1 : r_2 = \mu : a D.$$

Similarly, t is partitioned into the two type of twins in a ratio, given by

$$t_1 : t_2 = \Phi : b D,$$

whereas, s is wholly dizygotic.

Hence,

$$N_{12} = r_1 + t_1 = \frac{\mu r}{\mu + a D} + \frac{\Phi t}{\Phi + b D} \qquad (a)$$

$$N_{22} = r_2 + s + t_2 = \frac{a D r}{\mu + a D} + s + \frac{b D t}{\Phi + b D} \qquad (b)$$

Application of (*a*) and (*b*) however, requires the knowledge of a, μ, Φ and D.

The Differential Formula of Weinberg (1901):

The Weinberg's formula is deduced below, which involves the knowledge of r, s, t and the sex-ratio of single pregnancies. Now we have,

$$(i)\quad \frac{2s}{r+s+t} = \frac{4ab D(1-\mu)(1-\Phi)}{a\mu(1-\mu)^2 + b\Phi(1-\Phi)^2 + D(1-a\mu-b\Phi)^2}$$

$$(ii)\quad 1-\left(\frac{m-f}{m+f}\right)^2_{11} = \frac{4ab(1-\mu)(1-\Phi)}{(1-a\mu-b\Phi)^2}, \left[\because \left(\frac{m}{f}\right)_{11} = \frac{a(1-\mu)}{b(1-\Phi)}\right]$$

$$(iii)\quad \frac{N_{22}}{N_{12}+N_{22}} = \frac{D(1-a\mu-b\Phi)^2}{a\mu(1-\mu)^2 + b\Phi(1-\Phi)^2 + D(1-a\mu-b\Phi)^2},$$

from (15 and 16).

Combining (*i*), (*ii*) and (*iii*), therefore, we get

$$\frac{N_{22}}{N_{12}+N_{12}} = \frac{2s}{r+s+t} \cdot \left[\frac{(m+f)^2}{4mf}\right]_{11}.$$

Hence,

$$N_{22} = \frac{s(N_{11})^2}{2(mf)_{11}}, \quad \text{since } N_{12}+N_{22} = r+s+t.$$

This is a form of the Differential Rule, formulated by Bertillon (1874) and Weinberg (1901). The right hand side of the formula, deduced here agrees with the popular form; but it points out definitely that the sex-ratio to be used is appropriately that among the single confinements and not that among the total population as suggested by Weinberg. In practice, however, the difference will cause only very slight change in the result (Strandskov and Edelen, '46).

(20) Some interesting relations

$$(i)\quad \frac{r_1}{r_2} = \frac{\mu}{aD} = \frac{\text{Frequency of 1-egg } ♂♂ \text{ pairs}}{\text{Frequency of 2-egg } ♂♂ \text{ pairs}}.$$

(*ii*) $$\frac{t_1}{t_2} = \frac{\Phi}{b\,D} = \frac{\text{Frequency of 1-egg } ♀♀ \text{ pairs}}{\text{Frequency of 2-egg } ♀♀ \text{ pairs}}.$$

(*iii*) $$\frac{4\,r_2\,t_2}{s^2} = 1$$

(*iv*) $$r_2 : s : t_2 = m_{11}^2 : 2\,(m\,f)_{11} : f_{11}^2 .$$

(*v*) $$\frac{r_2}{s} = \frac{1}{2}\left(\frac{m}{f}\right)_{11}, \quad \frac{t_2}{s} = \frac{1}{2}\left(\frac{f}{m}\right)_{11},$$

$$\frac{r_1}{s} = \frac{\mu}{2\,a\,D}\left(\frac{m}{f}\right)_{11} \quad \text{and} \quad \frac{t_1}{s} = \frac{\Phi}{2\,b\,D}\cdot\left(\frac{f}{m}\right)_{11}.$$

(*vi*) Sex-ratio among monozygotic twins $= \frac{r_1}{t_1} = \frac{\mu\,b}{\Phi\,a}\cdot\left(\frac{m}{f}\right)_{11}^2$.

Sex-ratio among dizygotic twins $= \frac{r_2}{t_2} = \left(\frac{m}{f}\right)_{11}$, as before shewn in section 17.

(*vii*) As $$a\,D = \mu\,\frac{r_2}{r_1} \quad \text{and} \quad b\,D = \Phi\,\frac{t_2}{t_1},$$

$$D = a\,D + b\,D$$

$$= \frac{\mu\,r_2}{r_1} + \frac{\Phi\,t_2}{t_1}.$$

(21) Evaluation of μ, Φ, D and a in terms of quantities obtained from the births statistics:

In order to evaluate these four independent quantities, we require four independent equations connecting them. We have,

(*i*) $$a\,D = \frac{\mu\,r_2}{r_1}$$ from section 20 (*i*),

(*ii*) $$b\,D = \frac{\Phi\,t_2}{t_1}$$ from section 20 (*ii*),

(*iii*) $$\left(\frac{m}{f}\right)_{11} = \frac{a\,(1-\mu)}{b\,(1-\Phi)}$$ from section 14,

(*iv*) $$\frac{N_{22}}{N_{11}} = D\,(1 - a\,\mu - b\,\Phi)$$ from sections 14 and 15.

The quantities m_{11}, f_{11}, N_{11} and N_{22} are directly obtainable from the births statistics. Whereas r_1, r_2, t_1 and t_2 can be easily derived from the frequencies of the ♂♂, ♂♀ and ♀♀ pairs of twins.

We have from sec. 20 (*v*) :

$$\frac{r_2}{s} = \frac{1}{2}\left(\frac{m}{f}\right)_{11} \quad \text{and} \quad \frac{t_2}{s} = \frac{1}{2}\left(\frac{f}{m}\right)_{11}$$

$$\text{Therefore, } \frac{\text{frequency of 2-egg ♂♂ pairs}}{\text{frequency of ♂♀ pairs}} = \frac{1}{2}\left(\frac{m}{f}\right)_{11}$$

(*v*) Hence, frequency of 2-egg ♂♂ pairs

$$= \frac{1}{2}\left(\frac{m}{f}\right)_{11} \times \text{frequency of ♂♀ pairs.}$$

(*vi*) Therefore, frequency of 1-egg ♂♂ pairs
= (frequency of total ♂♂ pairs) $- \frac{1}{2}\left(\frac{m}{f}\right)_{11} \times$ (frequency of ♂♀ pair).

(*vii*) $\frac{r_2}{r_1} = \frac{\text{frequency of 2-egg ♂♂ pairs}}{\text{frequency of 1-egg ♂♂ pairs}}$, which is obtained by dividing (*v*) by (*vi*).

(*viii*) Frequency of 2-egg ♀♀ pairs

$$= \frac{1}{2}\left(\frac{f}{m}\right)_{11} \times \text{(frequency of ♂♀ pairs).}$$

(*ix*) Frequency of 1-egg ♀♀ pairs
= (frequency of total ♀♀ pairs) — (frequency of ♂♀ pairs) $\times \frac{1}{2}\left(\frac{f}{m}\right)_{11}$.

(*x*) Hence, $\frac{t_2}{t_1} = \frac{\text{frequency of 1-egg ♀♀ pairs}}{\text{frequency of 1-egg ♀♀ pairs}}$, which is obtained by dividing (*viii*) by (*ix*).

Now we obtain by componendo of (*iii*) and its reciprocal, the following equations :

(*xi*) $$(1 - a\mu - b\Phi) = a(1-\mu)\frac{N_{11}}{m_{11}} = b(1-\mu)\frac{N_{11}}{f_{11}}$$

From (*iv*) and (*xi*), we therefore obtain

$$\frac{N_{22}}{N_{11}} = D\,a(1-\mu)\frac{N_{11}}{m_{11}} = D\,b(1-\Phi)\frac{N_{11}}{f_{11}}$$

Substituting for $D\,a$ and $D\,b$ from (*i*) and (*ii*), we get

(*xii*) $$\frac{N_{22}}{(N_{11})^2} = \frac{\mu(1-\mu)r_2}{r_1 m_{11}} = \frac{\Phi(1-\Phi)t_2}{t_1 f_{11}}.$$

All the quantities except μ and Φ are known in the equations (*xii*), they can be solved for μ and Φ.

For the sake of brevity, we put

$$K_1 = \frac{N_{22}}{(N_{11})^2}\cdot\frac{r_1}{r_2}\cdot m_1 \quad \text{and} \quad K_2 = \frac{N_{22}}{(N_{11})^2}\cdot\frac{t_1}{t_2}\cdot f_1.$$

Therefore, $\mu(1-\mu) = K_1$, and $\Phi(1-\Phi) = K_2$.

(*xiii*) Hence,

$$\mu = \frac{1}{2}(1-\sqrt{1-4k_1}), \text{ and } \Phi = \frac{1}{2}(1-\sqrt{1-4k_2}).$$

The alternative roots of the equations are ruled out as μ and Φ are less than $\frac{1}{2}$.

(*xiv*) The quantity, D, can be evaluated by adding (*i*) and (*ii*). Thus,

$$D = \frac{\mu r_2}{r_1} + \frac{\Phi t_2}{t_1},$$

in which everything is known. D can also be determined from the relation, deduced in section 14. Thus,

$$1 - D = N_1/N\,(1 - a\mu - b\Phi) = q_1/(1 - a\mu - b\Phi).$$

(*xv*) The quantities, a and b, can be evaluated from the formula (*iii*). Thus, we have,

$$a/b = (m/f)_1 \frac{(1 - \Phi)}{(1 - \mu)},$$

and also remember that $a + b = 1$.

(*xvi*) Combining the expression for r_1, t_1, m_1, and f_1, deduced previously, it can easily be proved that,

$$\mu (1 - \mu) = r_1/m_1 = K_1, \quad \text{and} \quad \Phi (1 - \Phi) = t_1/f_1 = K_2,$$

since, we have in section 14, on neglecting δ and Δ,

$$m_1 = N a (1 - \mu) (1 - D), \quad \text{and} \quad f_1 = N b (1 - \Phi) (1 - D).$$

And from section 15, we have,

$$r_1 = N a \mu (1 - \mu)^2 (1 - D)$$

and

$$t_1 = N b \Phi (1 - \Phi)^2 (1 - D).$$

It should be kept in the mind that for the purpose of an ideal analysis, the confinement data must, first of all, be arranged according to the age-groups of the productive mothers. The computation of μ, Φ, D and a should then be undertaken for each of the age-groups of the mothers. Where such age-grouping is not permitted by the data, the analysis can still be applied to obtain informations about the plural birth characteristics of the population en masse. such an analysis is also promising to furnish much valuable results regarding the basic twinning tendencies of the different peoples. Two different ethnic variations are mixed up in such cases : one in respect of the twinning characteristics and the other in respect of the distribution of the mothers over the age-groups. The resultant effect is either an accentuation, or an obliteration of the differences, according as they act in the same or in the opposite directions.

(22) Summary

1. Almost all the statistical quantities needed for an elaborate expression of the twin data in a population have been expressed in terms of four basic probability parameters μ, Φ, D and a, explained in section 13.

2. Due considerations have throughout been made to limit the process of monozygotic scission or the process of dizygotic twinning at their appropriate numbers of occurrence so that a particular plurality of birth can take place. Essentiality of such limitations has hitherto been neglected (sections 2, 3, 15 and 16).

3. A differential formula has been deduced, which agrees with that given by Weinberg. It involves the use of the single birth sex-ratio in its application (section 19).

4. Formulae have been given to evaluate μ, Φ, D and a from the twin birth statistics classified according to the age-groups of the productive mothers concerned (section 21).

5. Certain refinements in the form of inclusion of the differential mortality of the two sexes, and the diminishing tendencies of scission with increase in the number of scissions of a zygote have been intentionally kept out of consideration for lack of sufficient and reliable data and partly for avoiding further complexities. Their inclusion is, however, felt essential for further improvements.

The author expresses his gratitude to Dr. B. S. Guha, the Director, Department of Anthropology, Government of India for his encouragement and to Dr. E. C. Büchi for constant help and suggestions.

Help from all quarters in the form of reliable twin and other plural births data will be gratefully received and acknowledged by the author.

REFERENCES

BERTILLON, L.-A., 1874. " Bull. Soc. Anthrop. ", Paris, 9 : 267-290.

CIOCCO, A., 1940. *Sex Differences in Morbidity and Mortality*, " Quarterly Review of Biol. ", 15 : 59-73, 192-210.

CREW, F.A.E., *Sex Determination*, 1946.

DANFORTH, C.H., 1916. *Is twinning hereditary ?* " J. Hered. ", 7.195-202.

DAHLBERG, G., 1926. *Twin births and twins from a hereditary point of view*. Stockholm, A.B. Tidens Tryckeri.

DAS, KEDAR NATH, 1934. *Twin pregnancy (A demographic and ethnic study)*. " Jr. Obs. Gyn. Brit. Emp. ", 41 : 227.

ENDERS, TRUDY and CURT STERN, 1948. *Frequencies of Twins relative to age of mothers in American populations*. " Genetics ", 33 : 263-272.

EVANS, H.M. and O. SWEZY, 1929. *The chromosomes in man, sex and somatic*. " Men. Mem. Univ. Calif. ", 9 : 1-65.

FISHER, R.A., 1919. *Genesis of twins*. " Genetics ", Vol. 4 : 489-499.

GALTON, F., 1875. *Short notes on heredity, etc. in twins*. " J. Anthrop. Inst. Great Brit. and Ireland ", 5 : 324-329.

GINI, C., 1923. *Di taluni studi recenti sopra le combinazioni sessuali e la frequenza dei parti plurimi e delle ricerche che essi suggeriscono*. " Giorn. di Biologia e Medicina sperimentale ", Vol. I, Fasc. 4.

GINI, C., 1924. *Sulle leggi della frequenza e delle combinazioni sessuali dei parti plurimi*. " Atti della Soc. Medico-Chirurgica di Padova ", 11 Mai 1923.

GREULICH, W.W., 1930. *The incidence of human multiple births*. " Am. Nat. ", LXIV : 142-153.

GREULICH, W.W., 1935. *Heredity in human twinning*. " Am. Jr. Phys. Anthrop. ", Vol. XIX (1934-1935) : 391-431.

HAMLETT, G.W.D., 1935. *Human twinning in the U.S. Racial frequencies, sex ratio and geographical variations*. " Genetics ", 20 : 250.

HELLIN, D., 1895. *Die Ursache der Multiparität der Unipaaran Tiere Ueberhaupt and der Zwilingsschwangerschaft beim Menschem*, Munchen.

JENKINS, R.L., 1927. *The interrelations of the frequency of plural births*. " J. Heredity ", Vol. 8 : 387 and 504.

JENKINS, R.L., 1929. *Twin and triplet birth ratios*. " J. Hered. ", XX, 485-494.

JENKINS, R.L. and JANE GWIN, 1940. *Twin and triplet births Ratios*. " J. Hered. ", 31 : 243.

KOMAI, T. and G. FUKOKA, 1936. *Frequency of multiple births among the Japanese and related people*. " Am. J. Phys. Anthrop. ", 21 : 433-447.

MORRILL, C.V., 1919. " Anat. Rec. ", 16 : 265-291.

NEWMAN, H.H., 1916. " Biol. Bull. ", 30 : 173-209.

OGUMA, K. and H. KIHARA, 1923. *Etudes des chromosomes chez l'Homme*. " Arch. de Biol. ", 33 : 493.

PAINTER, T., 1923. *The spermatogenesis in man*. " J. Exp. Zool. ", 37 : 291-334.

PAINTER, T., 1924. *The sex chromosomes of man.* " Am. Nat. ", 58 : 506-524.

PARKES, A. 2., 1926. *The physiological factors governing the proportions of sexes in man.* " Eug. Rev. ", 17 : 275-293.

PELLER, SIGISMUND, 1946. *A new rule for predicting the occurrence of multiple births.* " Am. J. Phys. Anthrop. ", 4 (n.s.) : 99-105.

PRICE, BRONSON, 1950. *Primary biases in twin studies* (*A review of prenatal and natal difference - producing factors in Monozygotic Pairs*), " Am. J. Human Genetics ", 2 : 293.

SARKAR, S.S., 1944-1945. *The frequency of Multiple births in India and Ceylon.* " Trans. Bose. Research Inst. " (1943-44), Vol. XV.

SCHULTZ, A.H., 1918. *Studies in the sex-ratio in Man.* " Biol. Bull. ".

STERN, C., 1949. *The principles of human Genetics.* W.H. Freeman & Co., California.

STRANDSKOV, H.H., 1945 (*a*). *Multiple birth confinement frequencies in the U.S. birth registration area from* 1922-1936 *inclusive.* " Am. J. Phys. Anthrop. ", Vol. 3 (n.s.) : 49.

STRANDSKOV, H.H., 1945 (*b*). *Sex-ratios of live-births and still-births of the U.S. total white and coloured population* 1922-36 *registration area.* " Am. J. Phys. Anthrop. ", Vol. 3 (n.s.) : 165.

STRANDSKOV, H.H. & H. BISACCIA, 1949. *The sex-ratio of human still-births at each month of Utero-gestation and at conception.* " Am. J. Phys. Anthrop. ", 7 (n.s.) : 131.

STRANDSKOV, H.H. & E.W. EDELEN, 1946. *Monozygotic and Dizygotic twin births frequencies in the White and the coloured U.S. Population.* " Genetics ", 31 : 483-446.

STRANDSKOV, H.H. & G.T. SIEMENS, 1946. *An Analysis of the sex-ratios among single and plural births in the total, the " white " and the " Coloured " U.S. Populations.* " Am. J. Phys. Anthrop. " (n.s.) 4 : 491-501.

TIETZE, C., 1948. *A note on the sex ratio of Abortions.* " Human Biol. ", 20 : 156-160.

WEINBERG, W., 1901. *Beiträge Zur Physiologie und Pathologie der Mehrlingsgeburten beim Menschen,* " Arch. f. ges. Physiol. ", 8 : 346-430.

WENTWORTH, E.N., 1914. *Sex in multiple births.* " Science ", Bd. 39.

WINNIVERTER, H.V., 1912. *Etudes sur la spermatogenese humaine,* " Archiv. Biol. ", 27 : 91-190.

ZELENY, C., 1921. *The relative numbers of twins and triplets.* " Science " (n.s.), 53 : 262-63.

G. HERDAN

Language in the Light of the Theory of Information

I. Introduction

Literary statistics or statistics of language usually meet with two rather important objections which, if true, would tend to make all efforts to apply statistical methods to language study futile. Before embarking upon a somewhat extensive statistical study of linguistic phenomena, it seems advisable to discuss these objections briefly because they may be at the back of the reader's mind.

1. *The objection as regards lack of randomness of samples.*

Samples of literary statistics as far as they deal with the frequency of occurrence of words, sounds, etc., are compiled from individual pieces of writing. Since the word material of any such literary product is dependent upon its subject, and is deliberately selected by the author, such pieces of writing cannot be regarded as random samples from a statistical universe or population. This would apply not only to the word material but also to its selection, or the method of withdrawing the sample.

2. *The objection as regards irrelevance of the statistical findings to the literary characteristics.*

A piece of writing is, in integrum, the outcome of a deliberately directed activity on the part of the author. Its dismember-

ment for statistical purposes into the number of words, be it according to their frequency of occurrence or according to their length in terms of number of syllables, or any other characteristic, destroys the character of the piece as a communication of thought, and whatever statistical relations we may discover in the dissected or "ground" sample lack relevance to the piece of writing as a whole.

ad 1. There is no such thing as a random method of selection per se, that is apart from the universe whose members are selected, and within the same universe, a method which is random in respect of one characteristic is not necessarily random in respect of another (1), (2). Kendall and Smith (3) therefore define a random method of selection applied to the characteristic *C* of the universe *U* as a method which is independent of *C* in U. In other words : selection will be random if the method by which the items, which are to form our sample, are selected, is uncorrelated with the characteristic we are sampling for.

Now the deliberate and individual choice of words in a piece of writing bears, as far as we know, no relation to the quantitative aspect of the word material, such as word length, no of sounds, etc. A word is chosen for its meaning or quality irrespective of its quantitative characteristics, except perhaps in poetry. Provided our statistical analysis of literary products is concerned with just these quantitative aspects, the piece of writing, whose material is collected by a method which is independent of the statistical feature in which we are interested, can be regarded as a random sample of those quantitative characteristcs of language. This is the theoretical explanation of why pieces of writing when dissected into the unit elements or formative elements of language may, in fact, show the characteristics of random samples from the universe of language.

This applies, however, in a stricter sense only to the universe of smallest or ultimate linguistic elements, the *phonems* (5). The more composite, on the other hand, the elements of the linguistic universe are, the more shall we be compelled to differentiate in a particulate language between universes according to the stage of linguistic development, the dialect, the subject of the piece of writing and even according to the individual writer. This applies, for instance, to the distribution of vocabulary and

to the distribution of word length in terms of the number of syllables per word.(1)

The interpretation given here to the emergence of laws of chance where they would be least expected to have validity is rather different from that sometimes found in the literature. Thus G. A. Miller (4) finds it "a bit surprising that one's own verbal behaviour follows such statistical rules with considerable orderliness. We select our words and arrange them to communicate with others with no consideration for relative frequencies of occurrence. Still, on the average, this apparently wilful and conscious behaviour of chosing words follows statistical rules with great regularity." The method adopted in this paper is to tune obedience to such laws of chance in a much flatter key. When expressing our thoughts by means of words, we are not only "apparently wilful and conscious," but really so, and that "We choose our words and arrange them with no consideration for relative frequencies of occurrence" of, say, the various phonems in the language, is just the reason for the emergence of laws of chance governing the distribution of the phonems.

ad 2. Such criticism is not quite overcome by pointing out that the phenomena of language like words, syllables, sounds, etc., can in fact be treated statistically like other material of the natural and social sciences, and that the statistics derived in that way have a right of their own to existence, irrespective of whether they satisfy the students of language and literature (5). However correct this argument is, it does not quite do justice to the fact that there is an important difference between the ways in which data are collected for the social or natural sciences and for purposes of language study. There we *collect* isolated observations which we form into our samples, and by means of which we describe the total; here we have a total before us which we *dissect* into what looks like formless elements, not in order to gain an integral picture of the data, though evidently with the purpose of learning a little more about them.

A powerful additional argument in favour of language statistics is this.

---

(1) In general, the universes of literary statistics are of the Lexis type, like most universes of social mass phenomena, where the conditions of simple random sampling do not apply.

There is one feature which all pieces of writing have in common: they confer information by signs or symbols. We shall see that to the symbols of any of the pieces of writing which we choose to investigate there apply the general characteristics which an efficient code for the communication of thought should possess according to the theory of information. The statistical distributions into which we dissolve such pieces of writing thus show the characteristic which we must suppose the whole to possess and to obey as its fundamental law. The dissection of the piece of writing into its constituent elements does, therefore, not destroy its essential feature, at least. The quality of the information has disappeared, but its quantitative features have been preserved.

What is most valuable in this is that *form* which is thought to belong to the piece of writing in integrum and which is believed to be destroyed by the statistical dismemberment, is seen to penetrate right down to the ultimate bricks of language, which so far were thought to be formless. This shows that the statistical treatment of language can produce results of interest to the linguist in his own sphere of study, and not only to somebody regarding language as just another object of the natural sciences.

Moreover, the statistical assemblies show not only the general characteristics of literary work, as means of communication of ideas, but, as will be seen, to some extent also their distinctive features, due to peculiarities of different authors in handling the materials of linguistics.

The following article is the first part of a more comprehensive study. It deals with the application of uni-variate information theory to language phenomena. In a continuation to it, translations from one language into another will be considered in the light of bivariate information theory.

II.

The Statistical theory of information.

The statistical theory which will here be applied to language is known as the theory of information or communication. It was developed by R. V. L. Hartley (6) and especially C. E. Shannon (7), originally for communication by telegraph and radio, and is, therefore, couched in terms familiar to telegraph and

radio engineers, but rather unfamiliar to either linguists or statisticians. A presentation of the theory from the standpoint of the mathematician is given by G. A. Barnard (8). The following brief presentation of its essential features will retain such of the original terms as are helpful for the intuitive understanding.

A communication system can be described schematically as follows. It consists essentially of five parts.

1. An information source which produces the message to be communicated to the receiving terminal. The message may be, for instance, a sequence of thoughts or of speech sounds.

2. A transmitter which operates on the message in some way so as to make it communicable. In spoken language it is the words or other linguistic units which transmit the sequence of ideas ; in written language it is the letters or other symbols which transmit the speech sounds. In telegraphy, the coding operation produces the sequence of dots, dashes and spaces corresponding to the message.

3. The Channel, that is the medium used for transmitting the message from the sender to the receiver.

4. The Receiver, which usually performs the inverse operation to that done by the transmitter, reconstructing the message from the signal. For instance, the decoding of the message sent by telegraph.

5. The Destination, that is the person or thing for whom or which the message is intended.

If we wish to consider the problem of coding in full generality, it is necessary to represent the various elements involved as mathematical entities, by suitably abstracting from their physical appearance. For the purpose of this paper, we need deal only with discrete systems of coding, that is those in which both message and signal are sequences of discrete units, a typical case of which is telegraphy.

Let the sequence of units used in the symbols of the channel or code be : $x_1, x_2 \ldots x_n$ where the subscripts 1, 2 ..... n stand for 1, 2..... n units, and let the transmittance of a given message contain n_1 occurrences of the symbol or symbols of length x_1, n_2 occurrences of the symbol or symbols of length x_2, and so on, the sum $n_1 + n_2 \ldots + n_n$ equalling N.

It seems reasonable from an intuitive idea of the function of a code that such a system, in order to be efficient, should use short and easily produceable symbols for frequently occurring parts of the message, and increasingly longer and less easily produceable ones for the less frequently occurring parts. In shorthand, for instance, the greater the frequency of a sound or sound combination, the more contracted and the easier to produce is the outline of the shorthand symbol.

To take an example from telegraphy. The post office provides four main types of special telegrams with standard or stereotyped text, — Congratulations, Birthday Greetings, Greetings, Best Wishes.

The symbols used for transmittance of such messages need, of course, only distinguish between the four types of stereotyped texts, and for reasons of economy the length of these symbols stands in inverse relation to their frequency. Thus (7).

Message	Probability	Code symbol	Length of Code
1. Congratulations	1/8	110	3
2. Birthday Greetings	1/8	111	3
3. Greetings	1/2	0	1
4. Best Wishes.	1/4	10	2

We may thus provisionally assume that in a good coding system the complexity of the symbols will be a function of the reciprocal of the frequency of their occurrence, or their probability p. In mathematical symbols, we write

$$x_i = f(1/n_i) = f(1/p_i). \qquad (1)$$

As regards the type of function, f, let us consider that if we wish to relate frequency of occurrence to length of symbol, the scales of both variable quantities should be made comparable. Since now frequencies change much faster than symbol lengths, and thus have a range of variability very different from that of symbol length, a logarithmic transformation appears indicated, and we write

$$x_i = K \log 1/p_i = -K \log p_i \quad \text{approx.} \qquad (2)$$

re where K is a constant depending upon the choice of base for the logarithmic function. (1)

The choice depends upon that of the unit for measuring information. Every actual event of information implies the decision between two possible symbols as the members of an alternative or opposition. It was found therefore that for many types of coding, the binary unit, that is one implying a simple or bilateral opposition, was the most appropriate or basic unit, and, consequently, that the log to the base 2 was the appropriate one to use. Arithmetically, the choice of the logarithmic base is a question of the scale and range of the variable symbol length. If that scale, as it does as a rule, starts at unity and continues by steps of unity, with the probabilities falling off fairly rapidly then the expression for x_i (formula 2) will be approximately true if the logarithm is to the base 2.

We thus write

$$x_i = -\log_2 p_i \text{ approx.} \qquad (2a)$$

and in terms of base 10, $x_i = -3.322 \log_{10} p_i$.

There exists an important relation between the mean symbol length of a message and the sum of all $-\log_2 p_i$.

If the quantity $\log_2 1/p_i$ is a good approximation to x_i, in which case the code may be regarded as efficient, then the following relation should be true (7)

$$\log_2 1/p_i \leq x_i < 1 + \log_2 1/p_i$$

Summing over the n classes of symbol length in the message and dividing by N, we get

$$\frac{1}{N} \Sigma n_i \log_2 \frac{1}{p_i} \leq \frac{1}{N} \Sigma n_i x_i < 1 + \frac{1}{N} \Sigma n_i \log_2 \frac{1}{p_i},$$

and writing H for $\frac{1}{N} \Sigma n_i \log_2 \frac{1}{p_i}$ and $\bar{x}$ for $\frac{1}{N} \Sigma n_i x_i$

we have

$$H \leq \bar{x} < 1 + H. \text{ (2)} \qquad (3)$$

---

(1) For a more stringent derivation of the basic formulae of information theory, in place of the intuitive one given here, see Sect. VII Part II of this paper.

(2) This is formula as given by BARNARD (8); SHANNON (7) has $1/N$ in place of 1.

In words: under conditions assumed to prevail in efficient coding or transmitting of information, the mean symbol length in messages of great length can never be less than H. This is the fundamental theorem of the theory of information.

For $p_1 = p_2 \ldots = p_n$, $H = -\log 1/n$, and is a maximum for a given n. The condition for H_{max} implies that there is no dependence whatsoever between symbol length and probability of occurrence, and H_{max}, therefore epitomises the most inefficient coding. The expression $1 - H/H_{max}$ or as we shall write $1 - H/H'$, is consequently, taken as a measure of the efficiency of coding for a given n.

Quantities of the form

$$\Sigma p_i \log_2 1/p_i = -\Sigma p_i \log_2 p_i = -K \Sigma p_i \log_a p_i \qquad (4)$$

where K is a positive constant depending upon the choice of the unit of measurement, and which in the case of a binary unit becomes 3.322, play a central part in information theory as measures of choice, information and uncertainty. Shannon recognised the form of H as that of Entropy as defined in certain formulations of statistical mechanicas, if p_i is the probability of a system of being in cell i of phase space, and he chose the symbol H for it because of Boltzmann's H-theorem. H is, therefore, called the entropy. If x is a chance variable, we will write for its entropy $H(x)$.

From Shannon's exposition, it would appear that he regards only H_{max} (for which we will write H') as characterising the state of complete randomisation or dissipation. But as the brief derivation of Boltzmann's law in the *Appendix* shows, H as well as H' represents a maximum of entropy in a given system, the latter the state of equidistribution of molecules in a volume of the system, the former the state of greatest probability of the system taking into account that the gas molecules possess different velocities or energies.

As shown in the *Appendix*, the difference between the two expressions for entropy H' and H, can be stated to be due to the former being calculated from the data with *one* subsidiary condition or constraint, viz. $\Sigma n_i = n$, whereas the latter under *two* constraints $\Sigma n_i = n$ and $\Sigma n_i \varepsilon_i = E$. Considering now that

$$\frac{\Sigma n_i \varepsilon_i}{n} = \bar{\varepsilon}$$

the second condition amounts to this, that the entropy is calculated under the condition of preserving the mean energy per molecule. But this is precisely how in the theory of information H is constructed [formula (3)]: the condition for efficient coding is such as to make $H = -\Sigma p_i \log_2 p_i$ approximate to $\bar{x}$ for great N. This shows H as representing the maximum entropy if the different lengths of words are taken into account.

III.

Language as an efficient code

1. Applying these ideas to language structure, we might argue on the following lines. Language is a system for the communication of thought by means of symbols, be it speech sounds (words) or written symbols. Since, for European languages at least, the spoken word is the fundamental or primary symbol, we shall in this paper concentrate upon the phonetic symbols as the means or "channel" for communication by language and since we are here interested in how the quantity of information is related to certain quantitative characteristics of the symbols of communication, we must fix our attention upon such characteristics of the linguistic symbols. The most prominent and general characteristics of code symbols are their duration upon which the speed of communication depends, and the facility of their production, upon which depends the convenience of their use. In the case of the spoken word, the symbol length may be expressed in terms of the number of syllables or of sounds and the facility of use of certain sounds by the number of combinations (according to the dictionary) into which these sounds enter. Of these two fundamental quantitative characteristics of linguistic symbols which may be investigated in connection with applying the theory of information to language, we shall deal in this paper with word length.

This first step in the quantitative understanding (in terms of no. of syllables in the spoken language) of language as a system of communication is admittedly an oversimplification of the actual position. The essential characteristic of coding symbols is that they possess distinctive value, i.e., that they enable different messages to be distinguished easily and cor-

rectly. While the number of syllables provides a first and fundamental token for distinguishing between words, (between words of different stem length, between stem and derivative words and between stem and flexions), the finer distinction between words of the same length is the work of the phonems of the language, that is such vowels and consonant sounds as have distinctive function. In spite of the violence which the restriction of the investigation to word length only, as different from syllable structure, imposes upon our feeling, it is the proper way to start an investigation of this kind.

We thus inquire whether the length of words in a given piece of writing stands to their frequency of occurrence in the relation required by the theory for efficient coding, and also whether different languages differ characteristically in this respect. In terms of the symbols of the theory of information : we shall calculate the quantity H for a given distribution of word occurrences according to the number of syllables ; ascertain whether it provides the lower limit to the mean number of syllables in that distribution, and compare the quantity $1 - H/H'$ for different pieces of writing in the same, and in different, languages.

The greater that quantity for a given language, the more would we feel justified in regarding that language as a system complying with the conditions for efficient coding. The property of the code which enables one to use the probability distribution of the symbols for making valid guesses at the message is called the *Redundancy* of the code (1). Since the magnitude of $1 - H/H'$ depends upon that property of the code, it serves as a measure of its Redundancy.

We shall, thus apply to distributions of word occurrences in pieces of writing the formula

$$H = -K \Sigma p_i \log p_i$$

and first compare these quantities with the corresponding mean length in terms of syllables.

---

(1) This is the definition given by W. MEYER-EPPLER, " *Informationstheorie*" *Die Naturwissenschaften, 39,* (1952) p. 341 which is in agreement with C. Shannon's use of the term. But the sense in which the term " redundancy " is used by different writers fluctuates.

TABLE 1. (1) — *Word Frequencies by Number (f) and per Mille (‰) in Samples from.*

Number of syllables	*Genesis* and *Exodus* incl. all words		*Genesis* and *Exodus*. excl. names etc.		*Shakespeare* Henry IV Pt. 2 prose		*Shakespeare* Henry IV Pt. 2 verse		*Bacon's Essays*	
(1)	(2) f	(2a) ‰	(3) f	(3a) ‰	(4) f	(4a) ‰	(5) f	(5a) ‰	(6) f	(6a) ‰
1	2,943	815	2,939	832	10,965	800	9,076	784	4,640	730
2	579	160	520	147	2,177	159	1,918	165	1,080	170
3	77	21	61	17	439	32	476	41	420	66
4	12	3	11	3	99	7	108		167	27
5	2	.6	2	.6	23		4	10	41	
6					2	2	—		3	7
7					2		—			1
Total	3,613		3,533		13,698		11,582		6,352	
Average number .	1.21		1.19		1.25		1.28		1.41	
Entropy *H*. . . .	0.81		0.75		0.91		0.96		1.22	

Numbers of syllables	*Gray's Poems.*		*Gray* (Letters)		*Johnson*		*Gibbon* 1st sample	*Gibbon* 2nd sample	*Gibbon* 1st and 2nd sample combined
(1)	(7) f	(7a) ‰	(8) f	(8a) ‰	(9) f	(9a) ‰	(10.1) f	(10.2) f	(10a) ‰
1	1,769	723	3,987	761	1,268	634	1,243	617	610
2	585	239	831	159	423	212	408	206	205
3	70	29	281	54	195	97	224	118	114
4	23	9	121	23	77	39	103	50	51
5	1	.4	15	3	29	14	21	6	9
6			2						
					8	4	1	3	1
Total	2,448		5,237		2,000		2,000	1,000	
Average number .	1.33		1.35		1.60		1.63	1.63	
Entropy *H* . . .	1.05		1.10		1.51				1.56 (combining both samples)

(1) Cols 2 to 12 are after Elderton (9).

Number of syllables	*Carlyle*					*Macaulay's Essays*			
	French Revolution	Past and Present	Heroes and Hero worship	Total		Clive	Hastings and Milton	Total	
	(11.1) f	(11.2) f	(11.3) f	(11) f	(11*a*) ‰	(12.1) f	(12.2) f	(12) f	(12*a*) ‰
1	713	692	811	2,216	696	1,260	723	1,983	662
2	180	193	208	581	182	375	189	564	189
3	92	73	88	253	80	201	95	296	99
4	32	34	39	105	32	67	42	109	37
5	8	12	5	25	8	14	17	31	10
6	2	3	—	5	2	5	2	7	3
7							1	1	3
Total	1,027	1,007	1,151	3,185		1,922	1,069	2,991	
Average number .	1.49	1.51	1.46	1.48		1.55	1.55	1.55	
Entropy *H* . . .				1.34				1.45	

Number of syllables	*G. B. Shaw* (Dramatic criticism).		*Words in Composite Sample of Modern English (Dewey).* (1) A.		*Words in Composite Sample of Modern English (Dewey).* (1) B.	
	(13) f	(13*a*) ‰	(14) f	(14*a*) ‰	(15) f	(15*a*) ‰
1	584	668	66,593	847	72,031	824
2	184	211	9,309	119	12,163	139
3	72	82	2,155	27	2,589	30
4	29	33	476	6	516	6
5 +	5	6	100	1	59	1
Total	874	1,000	78,633	1,000	87,358	1,000
Average number .	1.50		1.20		1.22	
Entropy *H* . . .	1.37		.76		.83	

(1) Of a sample of 100.000 word occurrences, only those occurring oftener than 10 times were recorded, and that in two ways: (*A*) every variant of a root was counted as a separate word; (*B*) all common variants of a root were grouped together as occurrences of one particular word.

Number of syllables	*Goethe* Wilhelm Meister		*Lichtenberg*		*Pushkin*		
	(16) f	(16*a*) ‰	(17) f	(17*a*) ‰	(18) f	(18*a*) ‰	
1	587	488	539	514	290	342	
2	410	342	317	303	307	362	
3	146	121	136	130	204	241	
4	49	41	49	46	39	46	
5	8	8	7	7	7	8	
6					1	1	
Total	1,200	1,000	1,048	1,000	848	1,000	
Average number .	1.74		1.73		2.02		
Entropy *H*. . . .	1.65		1.56		1.83		

Number of syllables	*Tolstoy* Autobiography		*Tolstoy* War and Peace (small sample)		*Tolstoy* War and Peace (large sample)		*Turgenieff* Rudin	
	(19) f	(19*a*) ‰	(20) f	(20*a*) ‰	(21) f	(21*a*) ‰	(22) f	(22*a*) ‰
1	393	331	174	310	371	336	334	328
2	369	311	189	336	356	322	293	288
3	253	214	117	208	220	198	241	237
4	120	100	53	93	107	96	123	121
5	37	31	22	41	40	36	19	19
6	10	9	5	12	9	9	6	6
7	5	4	1	—	3	3	1	1
Total	1,187	1,000	561	1,000	1,106	1,000	1,017	1,000
Average number .	2.20				2.21		2.24	
Entropy *H*. . . .	2.11				2.10		2.07	

TABLE 2. — *Statistics of the Distributions of Table* 1.

Col. No.	Source	M	H	Δ	H'	$1 - H/H'$
	English					
(2)	Bible, 1st Sample	1.21	.812	.40	2.32	.65
(3)	» 2nd »	1.19	.751	.44	2.32	.67
(4)	Shakespeare, Henry IV, prose.	1.25	.91	.34	2.32	.61
(5)	» » verse.	1.28	.96	.32	2.32	.59
(6)	Bacon.	1.41	1.22	.19	2.58	.53
(7)	Gray verse (total)	1.33	1.05	.28	2.32	.55
(8)	» prose	1.31	1.10	.21	2.32	.53
(9)	Johnson.	1.60	1.51	.09	2.58	.44
(10)	Gibbon	1.63	1.56	.07	2.58	.40
(11)	Carlyle	1.48	1.34	.14	2.58	.48
(12)	Macaulay	1.55	1.45	.10	2.58	.44
(13)	Shaw (Lit. Criticism).	1.50	1.37	.13	2.32	.41
(14)	Composite Sample from Modern English (*A*)	1.20	.76	.44	2.32	.67
(15)	Composite Sample from Modern English (*B*)	1.22	.83	.39		.64
	German					
(16)	Goethe (prose).	1.74	1.65	.09	2.32	.29
(17)	Lichtenberg	1.73	1.65	.08	2.32	.29
	Russian					
(18)	Pushkin, verse.	2.02	1.83	.19	2.58	.29
(19)	Turgenieff.	2.24	2.07	.17	2.71	.24
(20)	Tolstoy, I.	2.20	2.11	.09	2.71	.22
(21)	Tolstoy, II	2.21	2.10	.11	2.71	.22

It remains to choose the base of the log function, or the constant K.

Distinction between linguistic forms rests upon the logical opposition of two sound features with distinctive function, called *phonems*. This implies the bilateral opposition as the basic unit and, therefore, the log to the base 2 as the most appropriate.

Recalling also what has been said about this matter in Sect. II, the logarithmic base 2 would seem the appropriate one to use, since it would make the theoretical expression $-\log p_i$ a good approximation of the observed value x_i. We shall, therefore, apply formula (2*a*) for H as it is used in information theory generally.

2. Table 1 shows some distributions of word occurrences according to length in terms of the number of syllables for English, German and Russian.

Elderton (9) has discussed the doubts which beset the compiler of such data. There are, to be sure, cases where the classification of words according to the number of syllables may be doubtful. But, on the whole, there exists such agreement in the distributions from one and the same author, or between those belonging to one and the same epoch, and there are such differences between the distributions for different languages, overriding the much lesser ones between the distributions in one and the same language, that the method of measuring word-length by the number of syllables in spoken language, must be regarded as fairly consistent.

Discussion

We direct our attention first to the mean and H values. H is of the same order of magnitude but invariably less than M, as required by the theory for efficient coding. We shall see in Sect. V what it is that the closeness of agreement depends upon apart from the number N of observations.

The differences Δ are, strictly comparable only for the same number of degrees of freedom (classes of symbol length). In order to take this duly into account, we express each observed entropy H as a fraction of the entropy H', and form the quantity $1 - H/H'$. It expresses the efficiency of coding for the given distribution in such a way as to enable us to compare different distributions in this respect.

Table 2 lists, apart from M, H, Δ, H' also the quantity $1 - H/H'$. On the whole, we find that quantity higher for English than for German, and higher for German than for Russian.

Inspection of the value for Δ and $1 - H/H'$ shows that there is evidently a relation between these quantities ; they seem to be positively correlated. This is confirmed by the graph, fig. 1, where $1 - H/H'$ is plotted against Δ. The reason for this relation is not far to seek. If for a given M and number of degrees of freedom the gradient varies, H will vary, and in an efficient code that variation will be subject to the condition of the fundamental theorem. The steeper the gradient the smaller

will be H, and the greater, consequently, Δ, and also $1 - H/H'$. Thus Δ and $1 - H/H'$ should be positively correlated for efficient coding systems, the intensity of dependence increasing with the efficiency of coding.

What has emerged so far is that in the systems considered H never exceeds M which shows the distribution of words according to length in terms of the number of syllables to be systems of efficient coding. The relation of mean word length and H must be regarded as implying that the mechanism provided by language for conveying information is governed in its quantitative aspect by the principles which the theory of information has found to apply to efficient coding. (1)

In terms of the theory of information, we say that the entropy per symbol or per word occurrence is H units or "bits of information", the relative entropy, that is the entropy relative to H', is H/H' bits, and the efficiency of the code is measured in terms of bits of information by the quantity $1 - H/H'$. The greater that quantity, the more compressed is the series against that of equidistribution. Thus, for English we may say that the amount of compression is about 50 %, whereas for Russian it is only about 30%, and for German somewhere between.

Those with some knowledge of these languages will immediately appreciate that this expresses the different burden which the quantity of syllables puts on the speaker, as regards both physiological effort in producing the words and memory effort in retaining them. It is thus an index for the ease with which these languages are spoken by someone who could eliminate the mere habit factor. English is the easiest language in that respect, because the words are, on the whole, smaller and even a longish sequence of ideas can be expressed with comparatively little effort as regards production of sounds and memory. Much greater effort is required in both respects for speaking Russian.

The application of the theory of information has thus led to a quantitative expression, and assessment, of a commonly though only vaguely felt difference between languages. We also see

---

(1) In section V it is shown that the compliance with the fundamental theorem, though a necessary, is not a sufficient condition for the linguistic code of word length to be classed as 'efficient.'

from it that the statistical dismemberment of the pieces of literary work has by no means destroyed a difference which a reader would be aware of in perusing these works.

Pitman's Shorthand as an efficient code.

3. It was thought to be of interest to test the theory of information, in its main theorem at least, by applying it to a recognised system of efficient coding for the transmission of words, like Pitman shorthand. Whereas in the main we are concerned in this paper with the spoken language as a code for *thoughts*, it seemed worth while to consider a system which is concerned primarily with coding the spoken *word* and ease of writing. Shorthand is a method of writing so designed as to raise the aspect of speed to the rank of the governing principle. The more frequently a word occurs in speech, the shorter should its outline be in Shorthand; the oftener a sound or phonem occurs, the simpler should be its shorthand symbol. Pitman's shorthand is not orthographic in the sense that every letter of the word must be reproduced by a shorthand symbol, but phonetic; the same sound, no matter how it is spelt, has the same shorthand symbol. It may therefore be regarded as a code of the spoken word meant to replace the less efficient code of alphabetic writing.

In Pitman's shorthand, speed is attained primarily through the phonetic symbols being fewer in number than the letters in longhand orthographic writing, — whereas other systems like Gabelsberger shorthand achieve that purpose primarily by making their symbols tiny fragments of the letters, but writing a symbol for every letter of orthographic writing.

The principle of Pitman's shorthand is that a geometrical alternative or opposition is used to indicate a phonetic one. Thus, different directions and thickness of straight strokes \ | / are used to indicate different types of consonants; the alternative of direction for one and the same outline is used to indicate whether the consonant represented by the outline is followed, or is not followed by a vowel; the alternative between the right and left hook at the start of the outline is used to indicate whether *r* or *l* follows a consonant; the alternative between the right and left end hook to indicate whether a word closes

with an *n* or an *f*; the alternative between the forms) and o to indicate whether a vowel precedes the letter *s*; the doubling principle \ to indicate the derivative affix *-er*; the halving principle – to indicate the affix *-ed*; the large end hook for the syllable "tion" in two directions to indicate whether the preceding vowel is short or long; the device of a point put in front or at the end of the outline to indicate the prefix *con* or the ending *-ing*, respectively; the alternative between the three positions of the outline to indicate the type of the first vowel in the word.

Every outline in a piece of writing was examined with regard to the number of such alternatives or principles it contained, without counting that of position of outline since it occurs in every word, and thus would merely raise the count by unity in every outline. The following table shows the distribution of outlines of words — contractions of words were not included; — in two pieces of writing taken from a shorthand text of Goldsmith's Vicar of Wakefield.

TABLE 3. — *Distribution of Length of Pitman Shorthand Outline.*

Number of principles	*Shorthand* (Sample I)		*Shorthand* (Sample II)	
	f	‰	*f*	‰
1	503	478	445	431
2	260	247	264	256
3	143	136	136	132
4	100	94	108	104
5	41	39	56	54
6	6	6	19	18
			5	5
Total	1,053	1,000	1,033	1,000
Average number .	1.99		2.17	
Entropy *H* . . .	1.95		2.12	

The similarity of the distributions and of the parameters characterising them is evidence on the one hand for consistency in counting the number of principles, and on the other, for the total number of outlines from which the two samples were drawn to form a statistical collective in that respect.

The relation between M and H is that required by the fundamental theorem. Since we know Pitman's shorthand to be designed as an efficient code, we regard this as a confirmation of the correctness of the main theorem of efficient coding according to the theory of information. The compression H/H' is not very considerable, which is due to the gradient not being very steep. The correlation between Δ and $1 - H/H'$ is very clear here.

IV

Comparison of Samples of Word-occurrences

As was stressed in Sect. 1, the fact that a piece of writing consists of deliberately chosen words according to the subject and the individual style of the writer does not preclude its being a random sample with regard to certain language phenomena. If the method of selection is uncorrelated with such phenomena, then the piece of writing may yet be a random sample with regard to such phenomena. Can we regard our samples of word occurrences as random samples, if we arrange them according to word length? Since in deliberately choosing the words, the writer certainly does not consider the length of words in terms of number of syllables (except perhaps in poetry, and there only sporadically), there is at least no prima facie objection to such a view.

On the basis, however, of our empirical findings, we cannot escape the conclusion that the universe of word length in a given language can hardly be regarded as homogeneous. It appears that we may have to differentiate between collectives according to the stage of the linguistic development, according to dialects and even according to the individual writers.

Since, however, this question is of no great importance for the subject of this paper, we shall not attempt to decide it, and shall confine ourselves to the method of descriptive statistics in comparing different distributions of word length.

We first consider whether different pieces of writing of one author, and writings of different authors belonging to the same time, gave similar distribution of words according to the number of syllables. This is the case, e.g. with the different distributions

from Gray's poems which show a remarkable similarity. The same applies to the two samples from Gibbon and to those from Shakespeare, Bunyan and, in fact, in almost any case of twin samples.

That this is not, perhaps, due to the idiosyncrasy of the compiler of some of these samples (Elderton), or the similarity of subject, may be seen from the fact that the samples which I have collected from Russian writings show the same similarity among themselves. Two samples were drawn from works by Tolstoy, one from his Autobiography, and another from War and Peace. The first contained 1187 words, the second was compiled in two stages : first a sample of 561 words was collected and compared with the big sample from the Autobiography, then it was brought up to 1106 words and again compared. Reduced to 1000 these three samples show such similarity as one would hardly have supposed to prevail in samples from two such different pieces of writing, as the Autobiography dealing with the reminiscences of childhood, boyhood and youth, and War and Peace, dealing (in the section from which the sample was drawn) with Napoleon's invasion of Russia in 1812. The inference of a statistical universe as forming the source from which these samples were drawn as random samples receives support from the consideration of the fact that the difference in sample size appears to make no appreciable difference, provided N is large. The view which regards the samples from German as a distinct group receives support by an 11 million word count made by Kaeding in 1897 which gave 50% app. of monosyllables against the 48.8% and 51.4% of our samples (10).

Furthermore, the samples from one and the same language differ, on the whole, less among themselves than from language to language. There are, however, within a language characteristic systematic differences due to linguistic developments and due to the idiosyncrasies of the writer. In our English samples, word length is, on the whole, increasing with time which is surprising since it contradicts the general tendency towards monosyllabism in English with time. It finds its explanation in the fact observed by Yule (11) that the distribution function of vocabulary changes with time. The increasing richness of vocabulary has the consequence that the word occurrences tend to be less concentrated upon a few often used and, therefore, smaller words. This

in its turn produces an increase in the mean word length with time.

An interesting case is Turgenieff. Although his distribution of word length is quite characteristically different from those for English writers, and similar to samples from the other Russian sources, yet it resembles also to some extent the distributions which are typical for German, and thus expresses quantitatively the fact that Turgenieff, more perhaps than any other Russian writer of his day, was very much influenced by Western writing.

A distinction which one might be tempted to make from our material is that between literary and everyday language. Our list in Table 1 of samples from English contains apart from literary samples, in cols. 14,15 the parameters for a sample of about 80,000 words collected from modern newspapers, modern novels, etc., but excluding the literary classics of the 17th, 18th and early 19th century (10). The mean word length results considerably lower than for the examples from literary English prose. Considering "however" that not less than about a century has elapsed between these two groups, and that development of English tends towards monosyllabism, it is doubtful how much of the reduction is really due to difference in style and how much to linguistic development. There is yet another formal reason for the reduction in *M*. The compiler selected 100,000 words which he examined for numbers of occurrence. His final lists treats (*a*) every variant of a single root as a separate word ; for instance *large, larger, largest, largely*, were treated as four separate words. Of these he selected only those occurring more than ten times, obtaining thus a total of 78,633 occurrences with 1,027 different words. (*b*) As an alternative he grouped variants together, and obtained now 87.358 occurrences of root words with 1,131 different words. From these tables, I have calculated the mean length of words, and obtained 1.20 for the first, and 1.22 for the second. The increase, however slight, is, in my opinion, significant of the change in method, and affords also a clue to why the reduction in mean length is so considerable compared with the literary samples. In excluding all words with ten or less occurrences, a very much greater number of long words will be excluded, than of short words, because the longer the words, the smaller will, on the whole, be their

frequency. This must result in a depression of the mean value. In forming groups of variants, such long words with only few occurrences have a better chance of being included in the list, which accounts for the small increase in the mean value.

It follows from this that the difference in mean, and form of distribution, between the examples from literary and those from modern English can not be safely adduced as an argument in favour of the instability of samples of this nature.

A syllable count of spoken — as distinguished from written — English, in about 80,000 words occurring in telephone conversation gave a distribution of words astonishingly similar to that from Dewey table 1, cols. 14, 15. The frequency per 1,000 occurrences of words with 1, 2.. syllables resulted as follows ; (13)

Number of syllables	Frequency ‰
1	820
2	138
3	32
4	8
5 +	2

The fact that extensive records like those made by Dewey (12) and by French, Carter and Koenig (13) under so very different circumstances, lead to such similar distributions of word length is a further confirmation of the relative stability of the universe of word length in a given language at a given time.

The samples dealt with so far included all types of words. It might now seem reasonable to suppose that if we confined our sampling to one word category, say nouns only, we would change the property of the distribution and, with it the statistical parameters describing word length. In order to test this assumption, a count was made of nouns only in Bunyan's Pilgrim's Progress, Part I and II, and the following distributions according to word length resulted.

The means are 1.55 and 1.52 respectively, and thus, remarkably close to the overall mean of prose writers in English for whom we have obtained samples. This must be interpreted as an indication of the universe being insensitive to deliberate selection

according to certain word categories at least, which confirms its nature as a statistical collective.

TABLE 4. — *Distribution of Word Length of Nouns in Bunyan's Pilgrim's Progress.*

Number of syllables	Part I.		Part II.	
	f	%	*f*	%
1	2,429	600	2,451	610
2	1,129	279	1,105	276
3	372	92	383	95
4	101	25	73	18
5	16	4	4	1
Total	4,047	1,000	4,016	1,000
Average	1.550		1.524	
Entropy *H* . . .	1.44		1.38	

Finally it makes no appreciable difference to the distribution of word length, whether the word material represented a continuous piece of writing — (1st sample of Gibbon) or a collection of isolated words from these writings (2nd sample), whether the sample came from prose (1st sample of Shakespeare, Gray) or verse (2nd sample Shakespeare and Gray); also the exclusion of place names etc. made very little difference (2 samples from the Bible).

V.

THE CASE OF EQUALITY IN THE FUNDAMENTAL THEOREM.

During an investigation into the word length of Omar Khayam's Rubayat in Fitz Gerald's translation, it was observed that the frequencies of word occurrences with 1, 2, ... 5 syllables seemed to approximate closely to the terms of a geometric series with 2/9 as the constant factor. The series was (16)

x	1	2	3	4	5	Total
f	1832	420	88	12	5	2357
‰	777	178	38	5	2	1000
g. s.	783	174	38	8	1	

A similar close approximation, if not even better, is found for the sample from Shakespeare, also for that from Johnson, whereas other distributions, especially those from German and Russian samples do not show any such regularity. Table 5 gives the geometric series fitted to some of the distribution of Table 1.

The possibility of fitting such distributions by means of a geometric series was considered by Elderton to be no more than a fluke of chance (9). No significance was attributed to it, probably because one felt at a loss to account for such a regularity. Such an explanation, however, is provided in quite a striking way by the theory of information. We consider the fundamental theorem of that theory:

$$-\Sigma\, p \log p \leq M$$

For the sign of equality to apply, the $-\log p$ values must equal the variable, that is symbol lengths, excepting the possibility of exactly compensating deviations. Since the scale of the variable forms an arithmetic series, the $-\log p$ values must form such an identical series, and, consequently, the p's a geometric series.

Thus, for the variable 'word length' the series of the

$$-\log_2 p = \log_{10} p/\log_{10} 2 = -3.322 \log_{10} p$$

must be such that the differences between consecutive terms is unity.

For the 1st term we have

$$1 = -3.322\ (\text{Mantissa of } p_1 - 1)$$

or

$$\text{Mantissa of } p_1 = 2.322/3.322$$

which is satisfied for

$$p_1 = 0.5\ .$$

The mean of a geometric series (g. s.) is the reciprocal of its first term, and thus, in our case, 2.00, and the common factor as the difference between p_1 and unity is 0.5.

TABLE 5. — *Comparison with Geometric Series* (The column numbers refer to those of Table 1).

Number of syllables	*Grsy's letters*		*Carlyle*		*Macaulay*		*Johnson*		*Gibbon*		*Modern English (Dewey)*			
	col. (8)	g. p.	col. (11)	g. p.	col. (12)	g. p.	col. (9)	g. p.	col. (10)	g. p.	col. (14)	g. p.	col. (15)	g. s.
1	761	741	696	676	662	645	634	625	620	613	847	833	824	820
2	159	192	182	219	189	229	212	234	205	237	119	139	139	148
3	54	54	80	71	99	81	97	88	114	91	27	23	30	27
4	23	17	32	23	37	29	39	33	51	36	6	4	6	5
5	3	1	8	8	10	10	14	12	9	14	1	1	1	1
6 and over. . .	.3	—	2	3	3	.6	4	8	1	9				

Numbers of syllables Shortland Principles	*Goethe*		*Turgenieff*		Shorthand 1st sample		Shorthand 2nd sample	
	col. (16)	g. s.	col. (19)	g. s.	col. (3)	g. s.	col. (5)	g. s.
1	488	574	328	447	478	503	431	461
2	342	245	288	247	247	248	256	248
3	121	104	237	137	136	122	132	134
4	41	44	121	76	94	60	104	72
5	8	19	19	42	39	30	54	39
6		8	6	23	6	15	18	21
7			1	13		7	5	11
8								6

The whole series then becomes

Mantissa of p	p
2.322 : 3.322	.500
1.333 : »	.250
0.322 : »	.125
2.644 : »	.063
1.644 : »	.031
0.644 : »	.016
3.288 : »	.008
2.966 : »	.004
1.966 : »	.002
0.966 : »	.001
	1.0000

[Incidentally the geometric series is different from what is called Poisson's exponential function

$$P = \frac{e^{-m} m^x}{x!},$$

where the successive probabilities do not form a geometric progression but decrease according to the variable factor $\frac{m}{x}$].

This then represents the series of probabilities of symbol lengths which would result if they were calculated according to the basic theorem of the theory of information using the sign of equality in that theorem. The theory of information thus provides the explanation for the phenomenon in question, viz., the emergence, in certain cases of word length distributions, of a series of probabilities close to a geometric series. Of our samples, those from Pitman's shorthand provide the nearest approach to the ideal g. s. as described above.

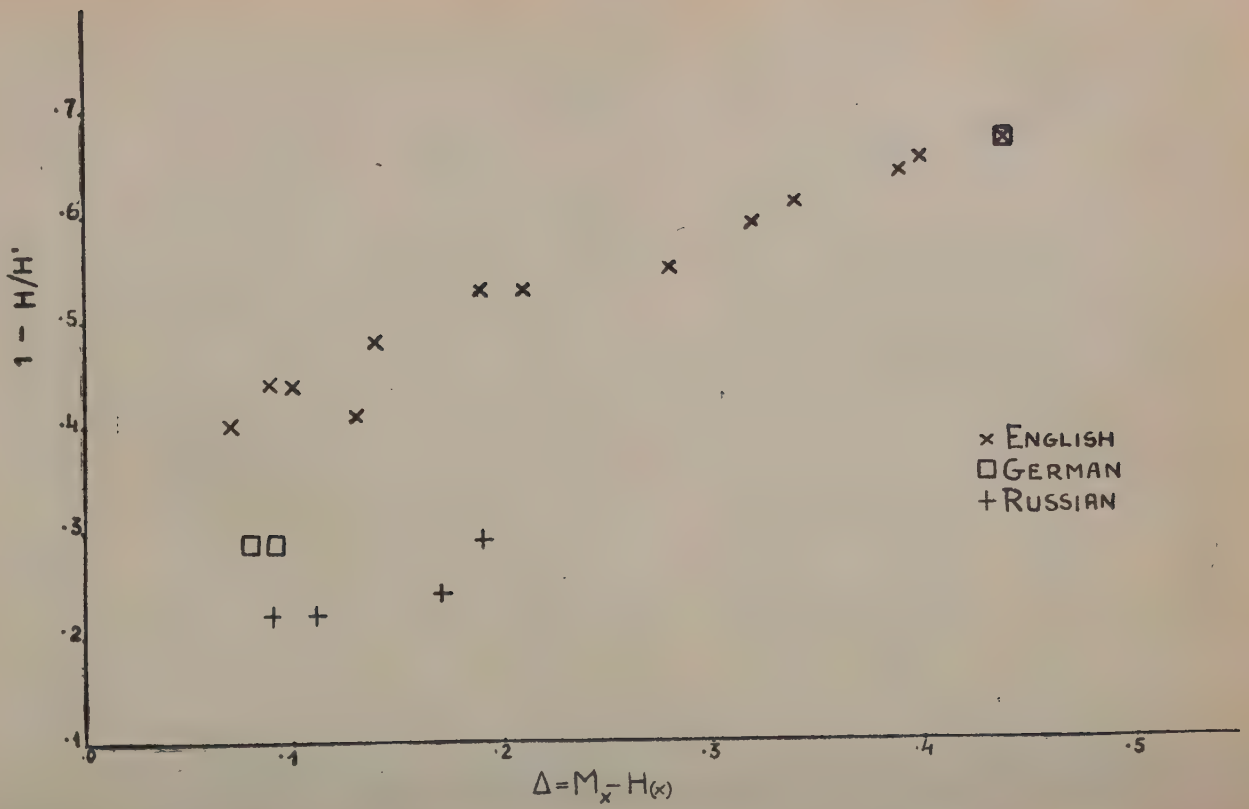
On the other hand, there are many word length distributions which do not approximate to such a series. In such distributions, H will approximate only more or less closely to M, according to the gradient of probabilities. It is only in the ideal case of the g. s. when $p_1 = .5$, that $M = H$ and $\Delta = O$. The more, on the other hand, the first term, p_1, exceeds .5, the smaller will be the common factor and the steeper the gradient, and the greater also $1 - H/H'$ and consequently Δ. As p_1 exceeds .8, the code

becomes inefficient. All provided that the observed distribution is a fair approximation to a g. s. The less the observed series of p conforms to a g.s., the less predictable becomes $1 - H/H'$ on the basis of Δ. But for the equality sign to apply and the relation to acquire the character of a functional relationship, it is not only required that the series of probabilities should be a g.s., but that it should be such a series with a specified first term, and, thus with a specified mean of the distribution.

This is well brought out by the graph, Fig. 1, where the English samples, which, to a greater or lesser extent, show a similarity with a g. s., show the dependence between Δ and $1 - H/H'$, whereas the German and Russian samples, being only poor approximations to a g. s. show a markedly smaller dependence between the variables. The difference Δ may be small, in such cases due to the compensating effect which non-systematic deviations from the ideal g. s. may have upon the mean M.

What can be inferred from all this is that redundancy, which depends merely upon the degree of compression of H against H', and thus upon the presence of a sort of gradient, determines the efficiency of coding not entirely by itself, but in conjunction with another element. This element is the property of the gradient to extend over a certain length, thus ensuring a fair range of variation among the symbols. This again depends, under the assumption of the gradient conforming sufficiently to a g. s., upon the probability of the first symbol in the series. Thus the efficiency of the linguistic code, whose symbols are words of different length in terms of syllables, appears to be the resultant vector of two component ones, viz., redundancy due to a gradient of probability so as to ensure sufficient differentiation for guessing word length ,and length of gradient so as to ensure sufficient range of the variable word length. The result will be a closer approximation of H to M.

Figure 1



Redundancy plotted against the difference between Mean and Entropy.

VI.

The Mechanism of the Linguistic Development towards Monosyllabism in the light of information theory.

So far, our interpretation of the distribution of word length was, so to speak, timeless. By this is meant that we regarded the source from which the samples were drawn as a stable universe, and all the elements in the sample as being obtained simultaneously. The result was that the word length seemed to be the determining factor for the frequency of use of words, the function varying according to language and, to some extent, also according to the writer: the shorter the word, the more frequent its occurrence. This is in accordance with the codes as we find them, e.g. for telegraphic communication, for shorthand writing, for certain post office communications. But strictly speaking, the position with these codes is somewhat different. It is not the length of the symbol which determines its probability of occurrence but, conversely, its probability of occurrence has been responsible for its length. A brief consideration will show that in language it is precisely the same, if we consider a distribution of words according to length as a product of time.

When comparing the samples from 18th and early 19th century writers and that from modern everyday English, reference was made to the change which word length undergoes in English with time as a result of the tendency towards monosyllabism. In respect of word length, the history of language moves along the vector of two forces: the agglutinating and the isolating forces. The former tends towards fusing root word and flexions into one whole, the latter towards isolation of syllables into independent words. Languages differ according to which of these forces is the stronger. Apart from the extremes of the two types of language, the Chinese as the almost perfect isolating, and the Indamerican languages as representatives of polysyllabic languages, most languages occupy an intermediate position in this respect. The European languages, for instance, as far as they belong to the Indo-Germanic family of languages, represent different stages towards monosyllabism. English may be said to be most advanced (Jespersen) (14), and the

Slav languages like Russian, least, with German and the Romance languages being of an intermediate type.

Not only do the different languages represent different stages of isolation, but within a language, like English, the progress towards isolation is a characteristic feature and powerful factor in its development. Viewed in this light, the universe of words, together with their probabilities of occurrence appears as the product of a process in time. A word with a great probability of occurrence, that is one which we find ourselves compelled to use very often, will be, with time, ground down to the smallest compass compatible with retaining its distinctive character, and thus become a monosyllable. The less frequent a word is used, the smaller the need for reducing its length. In this way, it is the necessity of use which makes language adjust the length of its symbols, just as in the case of devising an artificial code.

It follows that widely different periods of historical development of a language must, as regards word length, be conceived to be different universes. The same applies, in much stronger measure to different languages. The theory of information enables us to formulate and interpret the vaguely felt qualitative differences in a quantitative way, by making use of the differences between, say, English, German and Russian with respect to the descriptive parameters ; M, H, Δ, $1 - H/H'$.

These languages, as they represent different stages of development with regard to monosyllabism or isolation, so also different degrees of complying with the requirements of efficient coding. English, having the highest value of $1 - H/H'$ may be said to occupy the highest rank as an efficient code, German next, Russian least. Without wishing to be dogmatic on a point of linguistics which professional linguists may regard as too complicated to be decided on the slender evidence before us, we might tentatively suggest that the mechanism of the tendency towards monosyllabism may be sought in the principle of efficient coding, the gradient of probabilities for words of different utility being effective in progressively grinding down the length of words.

The term "*redundancy*" when applied to coding has its origin evidently in the fact that in an imperfect code there would be one code unit for every unit of message, that is in

our case, different word lenght for every word. As far as there is no such one-to-one correspondence, the code is regarded as redundant. Applied to our distributions, the smaller the variation in the number of syllable units which a language uses, per thought unit, — that is, roughly, per word — the more redundant is the linguistic expression, provided that the probabilities of different word length form a fairly systematic gradient. The more polysyllabic languages, like the Slav languages, having a fair range of word length but a rather poorly defined gradient, are therefore codes with a lesser degree of redundancy than are the more monosyllabic languages, which like English, possess a more systematic gradient of probabilities for different word length.

It might be thought that monosyllabism, carried to the bitter end, was the ideal coding for thought. But matters are not as simple as all that, as we know from Chinese. Although that language is essentially monosyllabic, yet the great number of written characters (upwards of 5,000 in a serviceable dictionary) and the comparative rarity of phonetic symbols (420 different syllables) make the language one of the most difficult to acquire and to understand, in brief, anything but an efficient code (15).

That complete monosyllabism is not an ideal state can be inferred from the theory of information by considering that in such a case there is practically no scale of symbol length and, therefore, no gradient of probabilities of symbols of different length. This represents the trivial case for which the fundamental theorem gives zero as the lower limit for the value of M. As the probability of monosyllables increases substantially above 0.8, H becomes less than unity, approaching zero as p_1 approaches unity. $H \rightarrow$ zero also becomes useless as a measure of the amount of information per word. Thus, the criterion for efficient coding, the basic theorem, loses its meaning. For the purpose of distinguishing between symbols, a systematic diversity of symbol length and of probabilities of occurrence of the classes of symbols is a requisite of efficient coding.

Yet another interesting inference leading to a better understanding of the mechanism at work, is provided by the theory of information when applied to our data.

Since efficiency in coding must be an important factor in the speed and accuracy of communication of ideas, the statistical theory of information would seem to lead to the unexpected

conclusion that languages differ in that respect. The commonly held view is that every language is more or less perfect in this respect for those who know how to use it.

That there should be an *objectively* determinable difference in the "goodness" of language is the new view we gain from applying the theory of information. Taking $1 - H/H'$ as the measure of "goodness" or efficiency of the language for the transmission of thought, English would, by this token, occupy top rank among the three languages examined, with German next and Russian last. It goes without saying that as we have so far examined only *one* aspect of the code, that ranking also refers to "goodness" only in that one respect.

Conclusions

1. Language in its quantitative aspect of word length fulfils the requirements of an efficient code as defined by the theory of information.

Different languages represent different stages of efficiency in coding the message of thought : English as the most advanced, with German next and Russian least.

2. The hitherto qualitatively known difference between languages as regards their approach to monosyllabism is replaced by quantitative assessment.

The theory of information provides an explanation of the mechanism of the development of languages in that respect with time.

3. The empirically found approximation of some of the distributions of word length to a geometric series is explained as the case of equality of the fundamental theorem for efficient coding.

4. The relation between H and H' of the theory is clarified by showing H to be the maximum entropy if differences in symbol length are taken into account as functions of the frequency of occurrence, whereas H' is the maximum entropy for the case of equidistribution, that is without letting different symbol length have any influence whatsoever upon frequency of occurrence.

BIBLIOGRAPHY

1. C. Gini, *Une application de la méthode représentative aux matériaux du dernier recensement de la population italienne*, « Bull. de l'Inst. de Statistique », Tome XXIII, 2e livraison, 1928, p. 198.
2. C. Gini, L. Galvani, *Di una applicazione del metodo rappresentativo all'ultimo censimento italiano della popolazione*. « Annali di Statistica » Series VI, Vol. IV, Rome, 1929.
3. M. G. Kendall and B. Babington Smith, *Randomness and Random Sampling Numbers*, J. Roy. Stat. Soc., Vol. CI, 1938.
4. G. A. Miller, *Language and Communication*. New York, 1951.
5. M. Boldrini, *Le statistiche letterarie e i fonemi elementari nella poesia*. Contributi del Laboratorio Statistico, Serie Sesta, dell'Università Cattolica del Sacro Cuore, Vol. XXI. Milano, 1948.
6. R. V. L. Hartley, *Transmission of Information*, Bell System Technical Journal, July, 1928, p. 535.
7. C. E. Shannon, *The Mathematical Theory of Communication*, The Bell System Technical Journal, Vol. XXVII, p. 379.
8. G. A. Barnard, *The Theory of Information*, J. Roy. Stat. Soc., Volume XIII, p. 46.
9. W. P. Elderton, *A Few Statistics on the Length of English Words*, J. Roy. Stat. Soc., Vol. CXII, p. 436.
10. F. W. Kaeding, *Haüfigkeitswörterbuch der Deutschen Sprache*. Berlin-Steglitz, 1897.
11. G. U. Yule, *A Statistical Study of Vocabulary*. Cambridge, 1944.
12. G. Dewey, *Relative Frequency of English Speech Sounds*. Harvard, 1923.
13. N. S. French, C. W. Carter, Jr., and W. Koenig, Jr., *The Words and Sounds of Telephone Conversations*. The Bell System Technical Journal, Vol. IX, p. 290, 1930.
14. O. Jespersen, *Growth and Structure of the English Language*. Leipzig, 1930.
15. B. Karlgren, *Sound and Symbol in Chinese*. London, 1923.
16. J. B. de W. Molony, *Private Communication by Letter*.
17. M. Born, *Atomic Physics*. London and Glasgow, 1938.
18. E. Schrödinger, *Statistical Thermo-Dynamics*. Cambridge, 1936.

Appendix.

The Boltzmann law of distribution is obtained as the most probable distribution of the particles of a gas in the various cells of arbitrary size into which we imagine the whole gas volume to be divided, subject to two subsidiary conditions,

$$\sum_1^z n_i = n \quad \text{and} \quad \sum_1^z n_i \varepsilon_i = E, \tag{5}$$

where the number of particles n and the total energy E are given.

1. To determine the number of molecules (of a gas) which, on the average, are to be found in a definite volume element we proceed as follows.

We imagine the whole volume of the gas, ω, to be divided into cells of varying sizes, $\omega_1, \omega_2 \ldots\ldots \omega_z$. The molecules are now imagined to be "thrown" at random into these cells; we then denote the number of molecules falling into cell ω_1 by n_1; the number falling into cell ω_2 by n_2, and so forth. The sum $n_2 + n_2 + \ldots\ldots n_z = n$, the total number of molecules in the volume W.

Denoting further the ratio of the size of cell ω_k to the total volume ω by

$$g_k = \frac{\omega_k}{\omega},$$

we have

$$g_1 + g_2 + \ldots\ldots g_z = 1.$$

Supposing that the total gas volume had been divided into two cells only of sizes ω_1 and $\omega - \omega_1$, respectively, the probability that of n particles in the total volume ω, the fraction n_1 will be found in the portion ω_1, is given by the Bernoullian law.

$$W = \frac{n!}{n_1!\,(n - n_1)!} \left(\frac{\omega_1}{\omega}\right)^{n_1} \left(\frac{\omega - \omega_1}{\omega}\right)^{n - n_1},$$

$$W = \frac{n!}{n!\,(n - n_1)!}\, g_1^{n_1} (1 - g_1)^{n - n_1}. \tag{6}$$

For z cells containing $n_1, n_2 \ldots\ldots n_z$ molecules respectively, we obtain for the probability of a definite distribution $n_1, n_2 \ldots n_z$ in the z cells by extending the above formula according to the polynomial theorem

$$W = \frac{n!}{n_1!\, n_2! \ldots n_z!}\, g_1^{n_1}\, g_2^{n_2} \ldots g_z^{n_z} . \qquad (7)$$

Transforming formula (7) for the probability of a definite distribution by means of Stirling's Theorems, which for large values of n gives

$$\log n! = n\,(\log n - 1),$$

we obtain for W, by taking logarithms,

$$\log W = \text{const.} + n_1 \log \frac{g_1}{n_1} + n_2 \log \frac{g_2}{n_2} + \ldots\ldots \qquad (8)$$

To find the most probable distribution, we calculate the maximum of $\log W$ for a variation of the numbers $n_1, n_2 \ldots\ldots n_z$ subject to the subsidiary condition

$$n_1 + n_2 + \ldots\ldots n_z = n .$$

We then obtain by the method of Lagrange Multipliers the equation for this maximum

$$\frac{\partial \log W}{\partial n_1} = \log \frac{g_1}{n_1} - 1 = \lambda .$$

$$\frac{\partial \log W}{\partial n_2} = \log \frac{g_2}{n_2} - 1 = \lambda , \qquad (9)$$

where λ is a constant.

It follows that

$$\frac{g_1}{n_1} = \frac{g_2}{n_2} = \ldots\ldots = e^{\lambda+1} = \text{const.} \qquad (10)$$

and

$$n_1 = g_1\, e^{-\lambda-1} .$$

Since now according to the subsidiary condition (5), the constant in the development is n, we have

$$n_1 = n\, g_1 = n \frac{W_1}{W}\,,$$

$$n_2 = n\, g_2 = n \frac{W_2}{W}\,. \tag{11}$$

We see that the numbers of molecules $n_1, n_2 \ldots n_z$ in the individual cells are proportional to the size g_z of the cells, — as was to be expected, — and that the distribution of the density of the molecules is, therefore, uniform. If all the z cells are of the same size, $\bar{g} = \frac{1}{z}$ and entropy defined as the average of the log of the reciprocal probabilities becomes

$$z \frac{1}{z} \log \frac{1}{\bar{g}} = \log \frac{1}{g}\,, \tag{12}$$

which has the form of H'.

2. For the distribution of velocities, we proceed in a similar manner. We construct a "velocity space" by drawing lines from a fixed point as origin representing as vectors the velocities of the individual particles in magnitude and direction. We then divide this "vector space" into cells and count the number of vectors whose ends fall into a definite cell.

But now we have a second subsidiary condition, viz.

$$n_1 \varepsilon_1 + n_2 \varepsilon_2 + \ldots\ldots n_z \varepsilon_z = E$$

where ε_i denotes the energy of a molecule whose vector points into the cell i, and where E represents the total energy of the gas.

Proceeding in the same way as before, we find for the maximum of the probability W

$$\frac{\partial \log W}{\partial n_k} = \log \frac{g_k}{n_k} - 1 = \lambda + \beta \varepsilon_k\,, \tag{13}$$

and

$$\frac{g_k}{n_k} = e^{1+\lambda+\beta\varepsilon_k}\,. \tag{13a}$$

This gives the Boltzmann Law of Distribution

$$n_k = g_k e^{-1-\lambda-\beta\varepsilon_k}$$

$$= g_k A e^{-\beta\varepsilon_k}. \qquad (14)$$

Writing

$$\frac{n_k}{g_k A} = p_k = e^{-\beta\varepsilon_k}.$$

we have

$$\log \frac{1}{p_k} = \beta \varepsilon_k,$$

or

$$\frac{1}{\beta} \log \frac{1}{p_k} = \varepsilon_k, \qquad (15)$$

which is of the same form as the expression for x_i in terms of p_i [formula (z)] for efficient coding.

The entropy σ is now defined as the average of the log of the reciprocal of the probability p, and thus as

$$\sigma = \overline{\log \frac{1}{p_k}} = -\Sigma\, p_k \log p_k, \qquad (16)$$

which shows it to be of the same form as H.

CORRADO GINI

Le basi statistiche delle diagnosi e delle prognosi (1)

In molti trattati di medicina vengono specificati, talvolta con minuti dettagli, i sintomi che accompagnano le varie forme morbose, ma come questi sintomi debbano venire combinati per eseguire la diagnosi non è detto, nè generalmente i trattati assurgono a considerazioni generali sull'attività diagnostica dei clinici.

Quei pochi che contengono considerazioni generali in proposito, distinguono una *diagnosi diretta* o *intuitiva* da una *diagnosi differenziale* che procede per via logica, ragionando sui sintomi e sulle loro combinazioni, e da una diagnosi, intermedia fra le due, detta *automatica* e, di fronte a queste, che vengono chiamate *diagnosi empiriche*, parlano di una *diagnosi razionale*, che non si limita a determinare la forma morbosa, ma ne ricostituisce altresì il decorso. Altri distinguono una diagnosi ottenuta seguendo il *metodo induttivo*, stabilendo, cioè, in totalità i singoli fenomeni del quadro morboso e traendo così la diagnosi dai fenomeni concreti, dalla diagnosi ottenuta seguendo il *metodo deduttivo*, quando si considerano solo alcuni sintomi individuabili con sicurezza e, dalla valutazione di questi, si traggono direttive di insieme, dalle quali successivamente, con piani più vasti, si potranno istituire ulteriori indagini speciali. Vi è chi pone fra le diagnosi ottenute col metodo deduttivo la *diagnosi per esclusione*, che viene accettata per vera dopo aver scartato, per la mancanza di uno o più sintomi fondamentali, tutti gli altri quadri morbosi tipici. Si distinguono infine una *diagnosi qualitativa* da una *quantitativa*, nella

(1) Conferenza letta il 9 giugno 1951 alle Riunioni di Statistica medica che facevano parte delle Riunioni medico-chirurgiche internazionali, tenute a Torino dall'8 al 12 giugno 1951.

quale ultima viene tenuto conto non solo della natura, ma anche della intensità dei sintomi. (1).

Non è mia intenzione di discutere tali classificazioni. Basta qui l'osservazione che in nessuna trattazione ho trovato specificato come questi sintomi, qualitativi o quantitativi, vengono, più o meno rapidamente e consapevolmente, combinati.

Persino Augusto Murri, che certamente ha fornito, tra noi, dei processi logici che devono dirigere l'attività del clinico, l'analisi più profonda, non va al di là di considerazioni generiche, rilevando alcuni pericoli, a cui nella diagnosi si va incontro, piuttosto che indicare il metodo che si deve in essa seguire e limitandosi, per la parte positiva, a insistere sulla necessità di fondere tutte le singole notizie in un giudizio sintetico, fusione in cui in sostanza si concreta il famoso *occhio medico*, mentre nei classici scritti di Claude Bernard — in quelli almeno che ho avuto modo di consultare — l'argomento non è neppure prospettato.

È stato perciò con vivo compiacimento che la Società Italiana di Statistica ha inteso, in una delle sue riunioni scientifiche, una relazione del compianto collega prof. Carlinfanti (2), nella quale la diagnosi (egli parlava di « diagnosi delle malattie infettive », ma le sue considerazioni si possono generalizzare a ogni forma morbosa) veniva ricondotta a una determinazione di probabilità delle cause. Se non che la stessa relazione del prof. Carlinfanti, per quanto pure pienamente fondata nell'idea direttiva, aveva avuto bisogno, come egli debitamente ricordava, di una nostra revisione critica delle formule proposte.

Ci è parso perciò interessante di rivolgerci a un certo numero di medici per sapere quali conoscenze sopra i sintomi essi ritengano necessarie per poter procedere alla diagnosi di una forma

---

(1) Vedi, in particolare, G. Viola, *Introduzione allo studio della medicina interna* (al paragrafo *La clinica come attività diagnostica*) nel trattato di *Medicina interna* diretto dal Prof. Angelo Cecconi, vol. I, Torino, Minerva Medica, 1936 e G. Pellegrini, *Delle Diagnosi*, nel trattato *Diagnosi medica differenziale*, diretto dal Prof. Adolfo Ferrata Vol. II, Wassermann, Milano 1939.

(2) E. Carlinfanti, *Probabilità delle cause e diagnosi delle malattie infettive*, discorso di chiusura della IX Riunione scientifica della Società Italiana di Statistica (Roma, 5-7 Gennaio 1949) in Atti della detta Riunione.

morbosa e come, nell'eseguire tale diagnosi, essi combinino i sintomi rilevati.

Solo pochi, appena una diecina, dei medici interrogati, hanno risposto, ed evidentemente le risposte sono venute da coloro che si ritenevano meglio in grado di darle, ma per verità nessuna fu soddisfacente.

Una non può dirsi errata, la quale dichiara che la probabilità, con cui, presentandosi un sintomo, si sia in presenza di una data forma morbosa, è uguale al rapporto fra i malati di detta forma morbosa che presentano il sintomo in parola e il totale dei malati che lo presentano. Se non che codesto rapporto, o meglio i due termini da cui esso risulta, non rappresentano in pratica il punto di partenza, ma il punto di arrivo del processo diagnostico, il quale consiste in una differenziazione tra le varie forme morbose, come esprime del resto la stessa parola « diagnosi », in quanto diagnosticare in greco significa distinguere, vale a dire differenziare una forma morbosa dalle altre.

Le altre risposte sono tutte difettose, in quanto tengono conto solo di una parte degli elementi che, in presenza di uno o più sintomi, sono necessari alla determinazione della probabilità di una forma morbosa.

Dovremo concludere da ciò che i medici, anche i più accurati e coscienziosi, seguono, nelle loro diagnosi, un procedimento logico errato ?

La conclusione sarebbe indebitamente pessimista. In verità, se ragionasse bene solo chi sa dire come si ragiona, l'umanità sarebbe composta nella sua strabocchevole maggioranza di persone che non sanno ragionare. È noto invece che si ragiona bene anche da persone che non hanno mai studiato logica, la quale effettivamente da pochissimi si studia e in molti paesi, fra i quali l'Italia, non forma materia di insegnamento, neppure sotto forma di corso libero, nell'ordine superiore, senza che perciò i cultori delle scienze vi restino al di sotto, per rigore di ragionamento, a quelli degli altri paesi.

Non vi parrà strano ciò che vi dico. Esso non è neppure nuovo. Già nel 600 i logici di Port Royal avevano definito come materia del loro studio « le riflessioni che gli uomini fanno sulle principali operazioni della propria mente », aggiungendo che « tutto ciò si fa naturalmente e talvolta meglio da quelli che non hanno imparato alcuna regola della logica che da quelli che tali

regole hanno imparato » e concludendo che « la logica non consiste nel trovare il mezzo di fare le dette operazioni, ma nel mettere in evidenza ciò che la natura ci fa fare ».

Molti fatti potrebbero addursi a sostegno di tale tesi. Fu già osservato, ad esempio, che Galileo, il quale, seppure non è stato, come fu affermato, il Cristoforo Colombo del metodo sperimentale, certamente ha dato di questo gli esempi più impressionanti, non ne ha però mai formulato una teoria neppure sommaria, mentre Newton, altra grandissima figura della scienza sperimentale, non è stato certo all'altezza del suo genio quando ne ha voluto codificare le norme. Se le sue celebri *Regulae philosophandi* fossero state osservate alla lettera, esse avrebbero invece ostacolato la ricerca scientifica. Fortunatamente Newton stesso fu il primo a non osservarle. *Hypotheses non fingo* egli, invero, proclamava, ma che cosa era l'idea della gravitazione universale, che ha reso il suo nome immortale, se non una tipica ipotesi che egli ha poi felicemente verificata ?

Instaurato dalla rivoluzione francese, il culto della Ragione ha avuto vita effimera nelle sue formalità esteriori, ma in realtà ha dominato e tuttora domina la scienza moderna.

Lo scienziato moderno ha continuato e continua a divinizzare la ragione quando la considera come qualcosa di primordiale ed immanente ed infallibile, superiore ad ogni altra manifestazione vitale e in particolare distaccata e anzi contrapposta all'istinto.

In realtà, quando si esamini la questione obiettivamente, la ragione appare, come era apparsa ai Giansenisti di Port Royal, come una facoltà innata della nostra mente, e quindi in largo senso istintiva, per quanto dagli altri istinti si distingua per la sua maggiore adattabilità alle condizioni esterne, una facoltà che, per quanto abbia raggiunto le sue massime espressioni nella specie umana, si è però radicata e gradualmente sviluppata negli esseri viventi attraverso una selezione non millenaria, ma milionaria, se non miliardaria, di anni.

Le operazioni dell'organismo umano si possono in verità distinguere in tre categorie :

1) Funzioni che si svolgono spontaneamente e, in condizioni normali, inavvertitamente e sono di conseguenza, nel loro decorso normale, incontrollabili. Tali la digestione e la circolazione.

2) Funzioni che si svolgono spontaneamente, ma vengono, volendo, anche nel loro decorso normale, avvertite, così che possono venire controllate. Tale la respirazione.

3) Funzioni acquisite, come il linguaggio e la scrittura, la marcia eretta ed il nuoto, talune delle quali, come il linguaggio e la posizione eretta, vengono acquisite da tutti gli esseri umani normali, mentre altre, come la scrittura ed il nuoto, solo da una parte di essi. Si acquisiscono tali funzioni o per imitazione o per istruzione, a volte per la combinazione dei due procedimenti.

La facoltà di ragionare rientra nella seconda o nella terza di tali categorie? Poichè non vi è dubbio che tale facoltà si possegga anche indipendentemente da ogni insegnamento e poichè, d'altra parte, trattandosi di una operazione interna alla nostra mente, mal si vede come potrebbe acquistarsi per imitazione, pare doversi concludere che essa appartiene alla seconda categoria, alla categoria, cioè, delle operazioni che si eseguono spontaneamente, ma di cui noi possiamo renderci conto e che possiamo pertanto controllare.

Gli scienziati hanno perduto occasioni eccellenti, che saranno sempre più difficili a verificarsi nel futuro, per eseguire ricerche decisive in proposito, studiando i bambini allevati dagli animali.

Tutti conoscono la storia di Romolo e Remo, allattati dalla lupa di Albalonga, ma tutti, o quasi, credo, la ritengono una favola, confortati in ciò dalla interpretazione che Tito Livio, tanto più vicino di noi all'avvenimento, ne ha riferito. Secondo tale interpretazione, Romolo e Remo sarebbero stati allevati da una donna che, perchè di mali costumi, veniva dai pastori chiamata « lupa » (1). Non sarebbe stata questa la sola volta che il nomignolo, preso da qualche animale da parte di una persona o di un gruppo di persone, lo fa scambiare con l'animale stesso. Quando nel 1928, in una visita alla Palestina, mi recai da Gerusalemme a Betlemme, e quindi al Mar Morto, la mia guida, giunti ad un dato punto del viaggio, mi ricordava che in quella località, secondo le Sacre Scritture, il profeta Elia vi sarebbe

(1) Il pastore del re, avrebbe trovato la lupa intenta ad allattare i gemelli: « Faustolo fuisse nomen ferunt: ab eo ad stabula Larentiae uxori educandos latos. Sunt, qui Larentiam, vulgato corpore, lupam inter pastores vocatam putent; inde locum fabulae ac miraculo datum » (*Titi Livii Historiarum, Liber primus*, Cap. IV).

vissuto, per un certo tempo, nutrito dai corvi. La leggenda — soggiunse il mio informatore — si spiega plausibilmente perchè effettivamente in questi paraggi si aggira tuttora una tribù di beduini che si denominano appunto « Corvi ».

L'interpretazione di Livio appare verosimile ed è accolta, per esempio, nella nota tragedia « Il solco quadrato » di Federico Valerio Giusto Ratti. Ma non sempre ciò che è verosimile è vero. A parte altri esempi leggendari, come quelli di Atalanta e di Mileto (1) da alcuni punti di vista analoghi al caso di Romolo e Remo, e quello di Zoroastro, che sarebbe pure statò allattato da una lupa nella cui tana, dopo averne distrutto la covata, alcuni malintenzionati l'avevano portato perchè ne fosse divorato — esempio che non potrebbe ricevere la stessa spiegazione data per il caso di Romolo e Remo — vi sono ormai numerosi esempi storici e contemporanei di bambini raccolti o rapiti ed allattati e poi allevati dalle lupe o dalle orse o da altri animali e non per pochi giorni, come sarebbe stato il caso di Atalanta, di Mileto, di Romolo e Remo e di Zoroastro, ma per anni, finchè vennero ritrovati e ricuperati dagli uomini.

Solo i casi ricordati per l'India salgono a ventisette — che forse sono da ridursi a ventitrè dopo l'eliminazione dei duplicati — nè appariranno molti, quando si consideri che verso il 1870 si calcolava da 5 a 6 mila il numero dei bambini che ogni anno vi venivano rapiti dai lupi. Basta ammettere che una su mille delle lupe, essendole morti i piccoli, fosse indotta, come nel caso di Zoroastro, ad adottare il bambino catturato perchè 5 o 6 bambini venissero allevati ogni anno in India. Supposta per essi una vita media di 4-5 anni almeno, il numero dei bambini conviventi in India coi lupi sarebbe stato normalmente di almeno 20-30.

Gli scrittori del tempo effettivamente affermano che era da ritenere che esistessero sempre in India bambini allevati dai lupi e ricuperati poi dagli uomini e, se la letteratura ne menziona complessivamente meno di una trentina, ciò è perchè è rimasto ricordo solo di quelli caduti sotto l'osservazione degli Europei a cui l'avvenimento sembrava strano.

---

(1) L'una e l'altro erano stati esposti da un genitore e raccolti da pastori; Atalanta era stata nutrita da un'orsa, Mileto da una lupa.

Non mancano, d'altronde, esempi di bambini allevati dai lupi anche per altri paesi e già Linneo (1) parlava di un caso verificatosi nell'Asia. Lo stesso Linneo citava esempi di bambini allevati da orsi, da pecore, da capre e da bovini (2). Per tutti questi animali, e particolarmente per gli orsi, altri casi si sono verificati in seguito ai quali qualcuno se ne è aggiunto per i porci, i leopardi, gli sciacalli, le scimmie.

Ricordava pure Linneo esempi di bambini che, abbandonati o dispersi o fuggiti da casa, avevano potuto sopravvivere coi loro propri mezzi ed erano stati ritrovati più tardi in condizioni selvagge ; e, anche di questi, altri casi si aggiunsero in seguito.

Degli uni e degli altri costituiva Linneo una particolare categoria zoologica che, sotto il nome di *Homo ferus*, contrapponeva all'*Homo sapiens*. Io credo che con maggiore proprietà si potrebbe parlare di *Homo dissociatus* in contrapposto all'*Homo socialis* (3).

E un terzo tipo di *Homo dissociatus* è fornito dai bambini che, per colpevole negligenza o crudeltà o delittuoso disegno di famigliari o di congiurati furono tenuti, durante lo sviluppo, isolati dall'umano consorzio e di cui anche oggi di quando in quando appare notizia sui quotidiani.

Discussioni e scetticismo sollevò l'affermata esistenza dello *Homo ferus*, particolarmente per ciò che riguarda il primo tipo di bambini allevati dagli animali, ma ogni scetticismo dovette cadere di fronte alla documentazione contenuta nel diario che il Rev. J.A.L. Singh, missionario indiano e Rettore dell'orfanatrofio di Midnapore, redasse e che successivamente fu pubblicato — insieme a un esposto particolareggiato degli altri casi di *Homo dissociatus* — dal prof. R.M. Zingg della Università di Denver (4)

---

(1) Nella 10ª edizione del suo *Systema naturae* (1758) e nelle successive.

(2) Camerarius, che per primo ricorda quest'ultimo esempio parla di *pecora*, parola latina di ampio significato, che qui probabilmente deve intendersi per *capre* e non per *buoi*.

(3) Vedi la mia conferenza dal titolo *Homo Socialis et Homo Dissociatus* letta il 13 settembre 1952 al XV Congresso Inter. di Sociologia di Istanbul e destinata ad essere pubblicata sugli Atti di detto Congresso. Una traduzione spagnola ne è frattanto apparsa sulla « Rivista Internacional de Sociologia ».

(4) Rev. J.A.L. Singh and R.M. Zingg, *Wolf-Children and Feral Man*, Harper Brothers, New York and London, 1939.

e riassunto dal prof. A. Gesell dell'Università di Yale, specialista di puericoltura di fama mondiale (1).

Si riferisce il diario a due bambine allevate dai lupi e ad essi sottratte alla presunta età di un anno e mezzo, l'una, e di 8 anni, l'altra, di cui la prima morì dopo pochi mesi, ma la seconda visse fino a 17 anni.

Ne risultano avvalorate — per le innegabili concordanze che il diario con esse presenta — anche le notizie relative ai casi precedentemente osservati.

Non tutti naturalmente sono ugualmente attendibili. Alcuni, anzi, sono più o meno gravemente sospetti.

Nè tutti possono fornire utili elementi di giudizio per la questione che qui ci occupa, sia perchè spesso i bambini dissociati vennero ben presto a morte, generalmente a causa della irrazionale pretesa dei loro educatori di imporre loro le abitudini di vita dell'uomo sociale, sia perchè, anche dei pochi che sopravvissero per qualche anno, mancano notizie sufficienti.

Ma particolarmente istruttivo è il caso di Kamala, la maggiore delle due bambine di Midnapore.

Al pari degli altri bambini allevati dagli animali, essa, appena strappata alla madre adottiva, era simile ad una bestia quanto una creatura umana può esserlo, abituata a camminare su quattro gambe e incapace di conservare la posizione eretta, avvezza a cibarsi esclusivamente di carne cruda e di carogne, della bocca servendosi direttamente per afferrare i cibi e per mangiare, come per bere alla guisa di un cane, intollerante di vegetali, di cibi cotti, di sale, intollerante del pari di vestiti, priva di linguaggio ed abituata ad ululare come i lupi, diffidente ed aggressiva verso i normali esemplari della specie umana e amichevole, invece, coi cani, e da essi ricambiata da spontanea fiducia, desiderosa di appartarsi in un angolo buio, attiva nella notte e a riposo nel giorno, dimostrando quasi una fotofobia, incline ad ansimare e tirar fuori la lingua quando era esposta al sole, mentre quasi non sudava.

Amorosamente ed intelligentemente educata dalla Signora Singh, abbandonò le abitudini ferine; perduta la propensione per le carogne, si abituò a nutrirsi di carne cotta e di legumi, a

---

(1) A. Gesell, *Wolf children and Human child*, Methuen and Co, London, 1941.

indossare vestiti fino al punto da rifiutarsi di lasciare il dormitorio e di uscire a passeggio se non era vestita come le sue compagne, a rispettare la proprietà altrui, a provvedere ai propri bisogni naturali secondo le esigenze sociali. Divenne paurosa, come gli altri bambini, del buio, e diffidente dei cani, che cercava di evitare. Apprese, a fatica, a camminare su due piedi, non però mai a correre, sempre usando per ciò dei quattro arti.

Quando, a 17 anni, venne a morte aveva imparato un centinaio di vocaboli con cui si esprimeva passabilmente bene, rendendosi pieno conto del loro significato e costruendo frasi di due o tre parole.

Non poneva questioni, ma rispondeva a tono ; sapeva esprimere i propri bisogni e i propri desideri anche servendosi di un gergo convenzionale. Cooperava con le persone a cui era affidata. Era divenuta da molti punti di vista un essere socievole, ragionevole e meritevole di fiducia. Prendeva iniziative, eseguiva ambasciate e cercava di rendersi utile con piccole incombenze. Rideva e piangeva, molto sensibile alle lodi che si meritava e scontenta degli insuccessi a cui andava incontro.

Aveva, in una parola, acquistata dignità umana. Se fosse vissuta, era da prevedersi che non avrebbe raggiunto un livello intellettuale sensibilmente inferiore a quello di altri membri della tribù primitiva da cui doveva essere sortita.

Più precoce nell'apprendere a parlare e a convivere si era dimostrata, probabilmente in ragione della sua meno lunga convivenza coi lupi, l'altra bambina, Amala, raccolta insieme a Kamala all'età di un anno e mezzo, ma morta poco dopo.

Un altro bambino, sottratto ai lupi in età di 5-6 anni e allevato pure in un orfanatrofio, non riuscì, forse per la minore abilità dei suoi educatori, a parlare, ma comprendeva quanto gli altri a gesti gli facevano intendere e a questi esponeva a gesti le sue richieste, per esempio per avere qualche soldo onde procurarsi da fumare [1].

---

[1] Vedi in proposito Jivani Jamshedij Modi, *Recorded Instances of Children Nourished by Wolfes and Birds*, (in « Asiatic Papers », Part II, Bombay, The Ternes Press, 1917, pagg. 197-200), dove sono contenuti maggiori dettagli che nel resoconto del prof. Zingg.

Questo articolo non è citato nella ricca bibliografia che chiude il volume di Singh e di Zingg. A complemento di detta bibliografia si possono pure citare un articolo dal titolo « *Wolf Children* » *Recorded Cases of a*

Dal diario del Rev. Singh plausibilmente conclude il Gesell che innata era in Kamala la capacità di sviluppare le facoltà mentali e di acquistare un livello culturale, facoltà che la convivenza coi lupi avevano bensì danneggiate ma non soppresse. Spontanee erano state le sue successive iniziative intellettuali e sociali e il conseguente progresso e non appreso dai suoi educatori anche se le intelligenti cure di questi potevano averlo facilitato. A volte — osserva il Gesell — si può dire che effettivamente fosse Kamala a insegnare ai suoi educatori, in quanto a questi essa dava a vedere una nuova capacità che da se stessa era venuta maturando.

Mentre è dunque ovvio che Kamala — del pari di Amala e dell'altro bambino sopra ricordato — ragionava, appare chiaro che le sue facoltà di ragionare erano innate. Difficile sarebbe stato d'altronde adattarsi altrimenti prima alla vita dei lupi e riadattarsi poi alla vita dell'umano consorzio.

Una certa facoltà di raziocinio non si può d'altronde negare, se non a tutti, almeno ad alcuni animali domestici, quali il cane ed il gatto.

Century pubblicato dal « Times of Ceylon » dell'8 maggio 1927 e riprodotto in « The Journal of Anthropological Society of Bombay (1927-28) », il volume di PER WESTERLUND, *Vargbarn*, Lindquist, Stockolm, 1947 e l'articolo di U. MARALDI : *Due bambine indiane allevate dai lupi in piena foresta*, su « Il Messaggero di Roma » del 2 dicembre 1951. Il Maraldi ha attinto a fonti di seconda mano e credeva che Amala e Kamala vivessero ancora, nel 1951, nell'orfanatrofio di Midnapore. Il Westerlund, che fu a Midnapore, ma non conosceva la pubblicazione del diario del Rev. Singh, riferisce per Kamala notizie che non sono tutte esatte. Nello stesso articolo del Maraldi si danno notizie, desunte dal «Sunday Express», di una bambina di una tribù Guayaqui che sarebbe stata raccolta nel 1932 all'età presumibile di 2 anni da una spedizione francese nella tana di una lupa a criniera che l'allattava e fu poi completamente assimilata alla civiltà occidentale, così che stava per prendere la laurea e per sposarsi con un bianco. Lo stesso « Messaggero », in un articolo del 2 ottobre 1952, dal titolo *Da una foresta del Giappone sono uscite due fanciulle scimmia*, parla di due ragazze di 15 e 17 anni, uscite dal bosco a saccheggiare un villaggio giapponese, che, dall'aspetto e dal comportamento, si presumevano vissute fin dall'infanzia fra le scimmie. Alcuni altri casi, più o meno attendibili, di bambini allevati da animali e non ricordati dallo Zingg sono citati nel discorso dal titolo *Wolf Children*, letto il 21 febbraio 1940 davanti alla « Folklore Society » dal suo presidente prof. J.H. HUTTON e riprodotto sulla Rivista « Folklore » vol. 41, marzo 1940, pagg. 9-31. Tutti meriterebbero accurato controllo.

I cani, ad esempio, per esprimere il desiderio di avere del cibo, assumono, da individuo ad individuo, attitudini completamente diverse, taluni di essi abbaiando, altri rizzandosi sugli arti posteriori, altri spiccando dei salti, altri appoggiando la testa dolcemente sulle gambe del padrone, altri invece spingendolo ruvidamente, ed altri ancora toccandolo con le zampe anteriori ed altri infine addirittura saltandogli addosso. Nè lo stesso animale fa sempre uso dello stesso comportamente con tutte le persone, ma sa convenientemente adattarlo, traendo partito dall'esperienza che ha fatto, alla persona ed alle circostanze con quel procedimento che viene ritenuto caratteristico del raziocinio. Noi usiamo dire in tal caso che all'animale non manca che la parola e ciò riprova che, per quanto le comunicazioni inter-individuali attraverso il linguaggio costituiscano un mezzo efficiente per lo sviluppo del raziocinio, tuttavia non sono una condizione essenziale per la sua esistenza. Coloro che credono ai cavalli parlanti e ai cani calcolatori, non avranno difficoltà ad ammetterlo.

Non meno istruttivo è il comportamento nei casi di *Homo dissociatus* del secondo tipo, rappresentato da bambini che, abbandonati o dispersi o fuggiti da casa, condussero vita selvaggia in condizioni di isolamento.

Poichè, anche in questo caso, i ragazzi trovati non posseggono linguaggio articolato o tutt'al più balbettano qualche parola, è chiaro che essi dovettero restare isolati fin dalla prima infanzia.

Se sanno invece, come in certi casi avviene, nuotare, arrampicarsi sugli alberi, usare di pietre e bastoni come armi di offesa e difesa, ciò significa che essi hanno appreso da sè tale comportamento, ciò che implica spontaneità di raziocinio. La stessa varietà, da caso a caso, del loro comportamento mostra quella adattabilità che caratterizza il comportamento razionale in contrapposto all'istintivo. Si è in presenza di vere e proprie reinvenzioni.

A volte essi hanno anche spontaneamente sviluppato attitudini che la comune degli uomini non possiede, quali quella di inseguire ed afferrare a nuoto i pesci, di passare da albero ad albero come animali arboricoli, di imitare per richiamo i suoni delle varie specie di uccelli. Qui trattasi di nuove invenzioni.

Assimilati — per lo più senza grandi difficoltà e rapidamente — al consorzio umano, essi hanno dato a volte prove di levatura mentale e di facoltà razionatrice non inferiori alle normali.

Tra i casi del terzo tipo di *Homo dissociatus*, rappresentati dai bambini a forza segregati dall'umano consorzio, fu a suo tempo famoso quello di Kaspar Hauser, secondo ogni verosimiglianza legittimo erede al trono di Baden, rapito nella culla per un complotto dinastico e rinchiuso fino a 17 anni in una fossa, dove giaceva seduto senza vedere faccia umana (neppure quella del suo custode) e senza altro passatempo che due cavallini di legno e qualche nastro per adornarli.

Liberato a 17 anni sapendo solo scrivere e ripetere macchinalmente il proprio nome e alcune frasi stereotipate e a mala pena camminare, dimostrò fin da principio un'acuta facoltà razionatrice oltre ad una memoria eccezionale e, nei quattro anni vissuti prima di cadere sotto il pugnale dei congiurati, aveva raggiunto il livello culturale normale in un adulto.

La spontaneità della facoltà razionatrice, che si deduce dagli esempi di *Homo dissociatus*, viene confermata dallo studio dei sordomuti. Anche quando non hanno appreso a comunicare mediante un linguaggio parlato o convenzionale, o prima che l'abbiano appreso, questi non sono affatto destituiti di raziocinio.

È istruttiva, a questo proposito, la storia di Elena Keller, la bambina che, divenuta cieca e sorda e muta a 19 mesi, potè, mercè l'opera di una istitutrice specializzata che ne assunse l'educazione a 7 anni, condurre una vita normale e, dal punto di vista intellettuale, anzi, molto superiore al normale, imparando, oltre alla propria, parecchie lingue moderne ed antiche, laureandosi a pieni voti a 24 anni, sapendo giocare a carte e a scacchi, praticando gli sports e lavorando, con successo, a migliorare le condizioni materiali e morali dei ciechi.

Il racconto della sua vita, che ella ci ha lasciato [1], è particolarmente interessante, per il nostro assunto, nella parte relativa al periodo che va dai 19 mesi ai 7 anni, in quanto esso ci mostra come, priva di comunicazioni col mondo esterno, le sue facoltà intellettuali si sviluppassero per forze interne, insofferenti di tale isolamento, che la successiva educazione le doveva permettere di rompere, estrinsecando le attitudini, già formate, di una mente normale.

---

[1] Elena Keller, *Il racconto della mia vita.* Traduzione di Clelia Allegri. Prefazione di Dino Provenzal. II Edizione, Carpignani e Zipoli, Firenze, 1923.

Tutto ciò costituisce, se non erro, una riprova della tesi sopra enunciata che la ragione umana rientra nella categoria seconda delle operazioni dell'organismo, delle operazioni, cioè, che vengono eseguite spontaneamente, ma possono essere avvertite e pertanto sono suscettibili di controllo. Ne segue che esse sono altresì suscettibili di perfezionamento. Come da varie parti si insiste nell'affermare che l'uomo deve imparare a ben respirare, chè, pur sapendo naturalmente respirare senza aver ricevuto alcun insegnamento, può da un insegnamento razionale trarre profitto, così apparirà naturale che l'uomo, pur sapendo d'istinto ragionare senza che ciò gli sia stato insegnato, possa dall'insegnamento avvantaggiarsi.

Non sta pertanto in contrasto con l'affermato carattere istintivo della ragione la insistenza con cui da molti si reclama un insegnamento della logica. Io sono tra questi. In sede di riforma universitaria, ho appunto proposto che l'insegnamento universitario si inizi con un anno propedeutico in cui siano impartite alcune materie, quali la matematica, la statistica e, in primo luogo, la logica, la cui conoscenza riesce ormai necessaria alla trattazione di tutti i problemi scientifici (1). Nè ero stato certo il primo ad invocare tale insegnamento. « Imparate a pensare » — ammoniva Federico Nietzsche in uno dei suoi ultimi volumi — « nelle nostre scuole non si ha più nessun concetto di ciò ;.... non il più lontano sentore che per pensare c'è bisogno di una tecnica, di un piano di istruzione, di una volontà di magistero — che il pensare vuol essere imparato, come vuol essere imparato il danzare, quasi una specie di danza » (2). Il paragone con la danza può essere discutibile, in quanto è discutibile se la danza costituisca una operazione acquisita (della categoria 3), come il linguaggio, o invece spontanea (della categoria 2), come abbiamo mostrato essere il raziocinio. Ma sul fondamento del monito non vi può essere disaccordo. Gli stessi logici di Port Royal, dopo di aver affermato il carattere istintivo del ragionamento, non negavano del resto che le riflessioni sul modo

---

(1) Vedi l'articolo *In tema di riforma universitaria*, « Civiltà fascista », anno X, n. 2, dicembre 1942.

(2) *Werke*, Leipzig 1906, B. X, pag. 291, citato da A. Gnudi e A. Vedrani nella *Prefazione* al volume *Pensieri e Precetti* di A. Murri citato a pagina 142.

con cui esso si svolge riescano in seguito di utilità. Senza tale convincimento, d'altronde, avrebbero essi mai insegnato la logica?

Che la capacità istintiva a ragionare, radicata nell'organismo da una selezione di milioni di anni, abbia per l'uomo bisogno di venire integrata con l'insegnamento per far fronte alle esigenze della moderna cultura, appare ben naturale quando si rifletta che tale selezione si è in realtà operata in base ad esigenze molto diverse da quelle che si incontrano nelle attuali ricerche scientifiche. Vi è diversità da due punti di vista essenziali. Da una parte, la logica istintiva riflette una situazione molto più stazionaria di quella che l'umanità attraversa all'ora attuale e senza confronto più stazionaria di quella a cui devono giornalmente far fronte, nelle loro ricerche, i cultori della scienza. Così come nei primi anni in cui si usavano gli aeroplani ancora a imperfetta tenuta d'aria, l'ufficiale di bordo ammaestrava i passeggeri sul modo di respirare per evitare le conseguenze dei rapidi mutamenti di pressione e come similmente si insegnano le precauzioni a cui, nel respirare, è bene attenersi di fronte all'impiego dei gas asfissianti o della bomba atomica, chè l'umanità non può essere stata ancora selezionata in base a tali nuove esigenze, così è necessario che i ricercatori vengano ammaestrati sulle precauzioni da adottare di fronte alle nuove situazioni che loro si presentano nelle ricerche scientifiche. Da tale punto di vista, la logica istintiva corrisponde, meglio che ad alcun'altra, alla condizione delle attuali società primitive, il cui studio approfondito potrebbe pertanto gettare molta luce sulla evoluzione dei nostri procedimenti logici.

D'altra parte, la logica istintiva si è formata essenzialmente in relazione a fenomeni individuali, mentre oggi la ricerca scientifica vede, nelle relazioni tra fenomeni individuali, l'espressione approssimata di più complicate relazioni tra fenomeni collettivi, solo suscettibili di essere determinate mediante l'ausilio della statistica. È pertanto precisamente in questo campo che i procedimenti logici istintivi hanno maggiore bisogno di controllo e di sviluppo.

Ora, il campo delle diagnosi mediche abbina le due difficoltà, quella del riferimento a fenomeni collettivi, che si rivelano non in un caso singolo ma in una massa di osservazioni, e quella di nuove contingenze a cui la selezione non può avere adattato il

raziocinio istintivo dell'uomo. Effettivamente i procedimenti, che il prof. Carlinfanti aveva presentato nella prima redazione, che ci aveva sottoposto, del suo lodato discorso, avevano avuto bisogno di venire integrati soprattutto quando la diagnosi, fondata sui sintomi che dapprima si erano offerti spontaneamente all'osservazione del medico, doveva poi essere integrata sulla base di altri, sopravvenuti successivamente, provocati con analisi chimiche, o medicine o altro trattamento clinico.

Ma è tempo ormai, dopo tanti preamboli, che spero non vi abbiano soverchiamente tediato, di venire alle formule, che spero pure non vi riusciranno troppo pesanti.

La probabilità, con cui un certo sintomo s, che un malato presenta, indica l'esistenza della forma morbosa a è, come sopra si è detto, data dal rapporto al numero delle volte N_s, in cui il sintomo s si presenta, del numero delle volte N_{sa} con cui s si presenta associato alla forma morbosa a.

Sarà cioè

$$ {}_s\pi_a = \frac{N_{sa}}{N_s} \qquad (1) $$

La formula vale sia che si tratti di un sintomo singolo che di un sintomo complesso, vale a dire di un insieme di sintomi (o, come a volte si dice, di una costellazione sintomatica). Salvo ulteriori precisazioni, parlando di sintomi in generale, intendo comprendere sia i sintomi singoli che i sintomi complessi.

Come si determinano i valori di N_{sa} e di N_s?

Se le forme morbose che possono presentare il sintomo s, sono in numero di n $(= a, b, c, \ldots n)$ il numero totale N_s delle volte in cui il sintomo si presenta sarà uguale alla somma delle volte in cui esso si presenta associato alla forma morbosa a, più il numero delle volte in cui esso si presenta associato alla forma morbosa b,..... più il numero delle volte in cui esso si presenta associato alla forma morbosa n; sarà, cioè,

$$ N_s = N_{sa} + N_{sb} + \ldots + N_{sn} \qquad (2) $$

D'altra parte, il numero delle volte N_{sa}, in cui il sintomo si presenta associato alla forma morbosa a, sarà dato dal numero delle volte che indicheremo con N_a, in cui si presenta detta forma morbosa a, moltiplicato per la frequenza che indi-

cheremo con ${}_aP_s$, con cui la forma morbosa *a* presenta il sintomo *s*. Sarà, cioè,

$$N_{sa} = N_a \times {}_aP_s$$

e analogamente

$$\left.\begin{aligned} N_{sb} &= N_b \times {}_bP_s \\ N_{sc} &= N_c \times {}_cP_s \\ &\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots \\ N_{sn} &= N_n \times {}_nP_s \end{aligned}\right\} \qquad (3)$$

Avremo quindi, in base alle (1), (2), (3)

$${}_s\pi_a = \frac{N_a \times {}_aP_s}{N_a \times {}_aP_s + N_b \times {}_bP_s + \ldots N_n \times {}_nP_s} \qquad (4)$$

la quale è appunto la nota formula della probabilità delle cause.

Essa ci dice che, a parità di altre condizioni,

1.) ${}_s\pi_a$ cresce col diminuire di *n*, ossia — come scriveva Augusto Murri — « il valore semiotico di un fenomeno [sintomo] è tanto maggiore quanto minore è il numero delle circostanze [forme morbose] capaci di suscitarlo » (1).

2.) ${}_s\pi_a$ cresce col crescere di ${}_aP_s$, e col diminuire di ${}_bP_s$, ${}_cP_s$, ... ${}_nP_s$, ossia — come scriveva Murri — « il valore [semiotico] di un singolo fenomeno [sintomo] non è [o, meglio, non è solo] in ragione della frequenza con cui si trova associato alla [forma morbosa considerata], ma piuttosto [o, meglio, anche] della rarità con cui si trova in malattie diverse da essa » (2). Questa giusta considerazione si trova fatta da tutti i medici che hanno risposto alla questione che loro avevo proposta (3).

---

(1) *Pensieri e Precetti* raccolti dalle sue opere per cura di A. Gnudi e di A. Vedrani, Bologna, Zanichelli, Nuova edizione, pagg. 205.

(2) Op. cit. pag. 206. Murri, in questo passo, parla in particolare della « tubercolosi del polmone » e non in generale della forma « morbosa considerata », ma l'argomentazione ha evidentemente carattere generale.

(3) Vedi, più sopra pagg. 128-129.

E tale unanimità di fondate risposte, inspirate al buon senso, è da rilevare in quanto statistici matematici e cultori di calcolo di probabilità, anche di grido, scambiano a volte ${}_aP_s$ per ${}_s\pi_a$ e credono di poter determinare questo avendo invece determinato quello con un errore di cui ho ripetutamente segnalato la portata (1). Onde non è da meravigliarsi se, anche nel campo dei medici, si incontri a volte tale errore proprio presso qualcuno che col calcolo delle probabilità e con la statistica ha una certa dimestichezza (2) confermando così che la bontà del ragionamento dipende più da intuizione che dagli studi compiuti.

Nè da Murri, nè — con una sola eccezione — dai medici da me interrogati, viene invece tenuto conto della circostanza, pure essenziale, che passiamo a considerare al numero seguente.

Si avverta, in particolare, che, per che sia ${}_s\pi_a = 1$, vale a dire perchè dalla frequenza di un sintomo s si possa risalire con certezza alla forma morbosa a, o, come si dice, per che il sintomo sia « patognomico », non è necessario nè sufficiente che sia ${}_aP_s = 1$, ma è necessario e sufficiente che sia ${}_bP_s = {}_cP_s = \ldots = {}_nP_s = 0$, vale a dire non è necessario nè sufficiente che il sintomo in questione si presenti sempre nella forma morbosa considerata, ma è necessario che esso non si presenti mai in un'altra forma morbosa.

3.) ${}_s\pi_a$ cresce col crescere di N_a e col diminuire di $N_b, N_c \ldots N_n$, vale a dire il valore semiotico del fenomeno (sintomo), per una forma morbosa, è in ragione della frequenza con cui si pre-

---

(1) Vedi particolarmente il discorso *I pericoli della Statistica* in « Atti della I Riunione scientifica della Società Italiana di Statistica », 9 ottobre 1939, riprodotto in « Rivista di Politica Economica », 1939, fasc. XI e la comunicazione *I testi di significatività* in « Atti della VII Riunione scientifica » di detta Società, 28-31 giugno 1943.

(2) Alludo al Tizzano, valoroso cultore di statistica medica e conoscitore sia dei miei passati scritti attinenti alla materia, sia della relazione del compianto Carlinfanti, il quale, dai sintomi della roseola, del tumore di milza e dalla reazione di Widal, vorrebbe dedurre la probabilità che si tratti di febbre tifoide, basandosi sulla frequenza con cui detti sintomi si presentano in questa malattia, senza tener conto della frequenza con cui essi si presentano in altre malattie. Cfr. A. Tizzano : *L'analisi statistica in biologia e in medicina*, conferenza tenuta alla riunione di Statistica medica in Torino, 9 giugno 1951, pubblicata nella « Rivista dell'Istituto Sieroterapico Italiano », Napoli, maggio-giugno 1952 e riprodotta nella rivista « L'Ospedale », Torino, settembre 1952.

senta la forma morbosa considerata e della rarità con cui si presentano le altre forme morbose.

Se, per esempio, una malattia *a* presenta un certo sintomo nell'1% dei casi, ma la malattia ha una frequenza di mille casi al mese, mentre una seconda malattia *b* presenta lo stesso sintomo nel cento per cento dei casi, ma ha una frequenza di 1 al mese, di 11 casi che si presenteranno in un mese, 10 apparterranno alla malattia *a* e solo uno alla malattia *b* per quanto quest'ultima presenti il sintomo in questione tanto più spesso della prima.

Tener conto della diversità dei valori N_a, N_b, ..., N_n è, ai fini della determinazione di ${}_s\pi_a$, altrettanto importante che tener conto della diversità dei valori ${}_aP_s$, ${}_bP_s$, ... ${}_nP_s$, in quanto quelli e questi entrano simmetricamente nella formula (4).

È perciò singolare che i medici, nel teorizzare intorno alla diagnosi, portino la loro attenzione con tanta preferenza sulla diversità dei valori di ${}_aP_s$, ${}_bP_s$, ... , ${}_nP_s$ e trascurino invece le diversità dei valori N_a, N_b, ... N_s. Ma ciò non importa necessariamente che, nel fare praticamente le loro diagnosi, i medici prescindano dalle diversità dei valori di N_a, N_b, ... N_n e, se, ad esempio, due malattie *c* e *d* presentano lo stesso sintomo con la stessa frequenza (essendo, cioè, ${}_cP_s = {}_dP_s$), ritengano, quando il sintomo *s* si presenta, di poter essere con la stessa probabilità in presenza della malattia *c* o in presenza della malattia *d*, anche se la malattia *a* è rarissima e la malattia *d* frequentissima. Così non è certamente. Non vi è dubbio che i medici tengano in pratica conto anche della diversità di frequenza delle due malattie, ma lo fanno istintivamente, senza rendersene conto e senza renderne conto a chi — come io ho fatto — pone loro la questione.

4.) Un'altra cosa — importantissima — la (4) ci dice: essa ci dice come i valori N_a, N_b, ... N_n, ed i valori ${}_aP_s$, ${}_bP_s$, ... ${}_nP_s$ si debbono combinare ai fini della determinazione della probabilità cercata ${}_s\pi_a$. Pur tenendone conto secondo le norme esposte ai numeri 1), 2) e 3), tali valori potrebbero combinarsi in modi molto diversi che a tutta prima potrebbero anche parere plausibili. Per esempio, la probabilità *a priori* (prima, cioè, di avere osservato il sintomo *s*) di essere in presenza della malattia *a*, è $V_a = \frac{N_a}{N_a + N_b + ... + N_n}$; la probabilità *a posteriori*

(dopo, cioè di avere osservato il sintomo *s*), nell'ipotesi che la probabilità a *priori* fosse uguale per tutte le malattie, è

$$\tau_a = \frac{{}_aP_s}{{}_aP_s + {}_bP_s + \ldots + {}_nP_s}$$

È certo che ${}_s\pi_a$ cresce con V_a e con τ_a. Se di V_a e τ_a noi facessimo la media aritmetica o il prodotto, il risultato terrebbe conto sia dei diversi valori $N_a, N_b, \ldots N_n$ che dei diversi valori ${}_aP_s, {}_bP_s, \ldots {}_nP_s$, in perfetta armonia con quanto è stato indicato ai numeri 1), 2) e 3), ma non darebbe affatto (salvo casi particolari) il valore esatto di ${}_s\pi_a$ che si desume dalla formula (4).

Ora, nè Murri nè alcuno dei medici che hanno risposto alla mia inchiesta ha saputo dire come egli combinava o combina i vari elementi di giudizio della diagnosi, che pure certamente, per arrivare alla diagnosi, in qualche modo doveva o deve combinare. Murri ripetutamente insiste sulla necessità di fondere tutti gli elementi di giudizio, ma come si debbano fondere, come egli stesso nelle sue diagnosi li fondesse, non dice. Evidentemente egli non se ne rendeva conto. Finissimo analizzatore del ragionamento clinico — egli pure ragionava di istinto, in parte inconsciamente.

Ma se dunque — si dirà — nelle diagnosi loro i medici ragionano bene d'istinto, di che preoccuparsi ? Lasciamo che l'istinto operi, senza intralciare i provvidi meccanismi della natura.

Se non che, per bene diagnosticare, non basta ragionar bene, bisogna anche ragionare su dati esatti ; non basta sapere che si deve tener conto dei valori $N_a, N_b, \ldots, N_n$ e dei valori ${}_aP_s, {}_bP_s, \ldots, {}_nP_s$ e come se ne deve tener conto ; bisogna anche conoscere quali sono i loro valori esatti. Ora, ciò non si può fare con l'osservazione comune, poichè la nostra mente, anche nelle persone più intelligenti e dotate di più forte memoria, è incapace di sintetizzare le relazioni numeriche concernenti i fenomeni collettivi. Di qui, in generale, la necessità, ai fini delle ricerche scientifiche, delle rilevazioni statistiche ; di qui, in particolare, la necessità, ai fini di una diagnostica scientifica, di statistiche che permettano di determinare i valori di $N_a, N_b, \ldots, N_n$ e di ${}_aP_s, {}_bP_s, \ldots, {}_nP_s$ e quindi di $N_{sa}, N_{sb}, \ldots N_{sn}$ e infine di ${}_s\pi_a$.

Sarebbe perciò desiderabile che ogni medico, chiamato a curare o constatare una data forma morbosa, compilasse una

scheda da cui, accanto alla forma morbosa diagnosticata e, se del caso, a quella in definitiva accertata, figurassero i sintomi qualitativi e quantitativi che ha presentato. Le schede compilate potrebbero venire raccolte ed elaborate da un apposito ufficio.

In pratica, converrà accontentarsi, in un primo tempo, di far compilare tale scheda nelle case di cura e negli ospedali, salvo ad accordarsi tra questi al fine di mettere insieme tutte le schede raccolte, in modo da raggiungere più presto il numero di casi sufficienti a conferire attendibilità ai risultati dello spoglio.

S'intende che in tali schede, e quindi nei risultati che se ne ricavano, non si potrà tener conto di tutte le modalità e gradazioni dei sintomi, le quali pure ad un medico esperto possono riuscire a volte decisive per la formulazione della diagnosi, cosicchè l'*occhio clinico* avrà sempre occasione di esercitarsi. Il medico troverà però nelle statistiche fornitegli dallo spoglio, una guida, sia pure sommaria, ma in ogni caso molto più sicura di quanto non possano essere oggi le sue impressioni, desunte dall'esperienza personale o dalla letteratura, sulla frequenza delle varie malattie e, in ogni malattia, dei vari sintomi.

Una guida siffatta sembra essere indispensabile per i clinici chiamati a diagnosticare fuori delle loro sedi, in quanto da località a località può variare, e a volte decisamente variare, la frequenza delle varie forme morbose, per modo che chi, sia pure con perfetto rigore di metodo, diagnostichi sulle forme morbose di un paese *A* sulla base della frequenza delle malattie (vale a dire dei valori N_a, N_b, ... N_n) ricavati dalla sua esperienza fatta in una località *B*, può incorrere in errori grossolani.

Si potrà osservare che, ai fini del trattamento delle forme morbose, non è necessario conoscere esattamente, e in molti casi neppure approssimativamente, il valore di ${}_s\pi_a$ (la probabilità, cioè, che ci si trovi in presenza di una data forma morbosa); basta conoscere qual'è la forma morbosa più probabile, conoscenza che spesso si può ricavare dalla propria esperienza personale, senza ricorrere a rilevazioni statistiche lunghe e costose.

Se non che ciò non è vero, o per lo meno non è vero sempre. È vero solo quando le forme morbose, tra cui conviene differenziare, hanno la stessa importanza per la vita o la salute o la borsa del paziente; ma, se un medico trova un sintomo che gli fa pensare che col 10% di probabilità è in presenza di una malattia gravissima *a* e col 90% di probabilità in presenza di

una malattia *b* di pochissimo rilievo, evidentemente dovrà curare il paziente in base all'ipotesi della malattia *a*, non in base alla ipotesi contraria, ammesso che i due trattamenti si escludano a vicenda. Se non che per tale giudizio gli è necessario conoscere i valori di ${}_s\pi_a$ e di ${}_s\pi_b$ (nel nostro caso 0,1 e 0,9). Se invero, la probabilità di trovarsi in presenza della malattia *a*, invece che essere del 10% fosse solo dell'1% o del ½%, per quanto la malattia *a* sia più grave della malattia alternativa *b*, potrebbe essere conveniente trattare il paziente per la malattia *b*.

Comunque rilevazioni statistiche sistematiche si rendono sempre necessarie quando si è in presenza, non di sintomi singoli, ciò che è del tutto eccezionale, ma di costellazioni sintomatiche più o meno complesse, perchè, se, in base all'esperienza personale, si può avere un'impressione più o meno approssimata, ma in ogni modo attendibile, della frequenza con cui un singolo sintomo si presenta in varie forme morbose, ciò è praticamente impossibile per costellazioni sintomatiche appena un poco complesse. Si potrebbe — è vero — pensare a risalire dalla frequenza dei singoli sintomi a quella delle loro combinazioni, ma è impresa praticamente inattuabile, in quanto i vari sintomi non solo di regola tra loro nè equipollenti nè indipendenti, cosicchè il calcolo, in base all'ipotesi che siano equipollenti ed indipendenti, porterebbe a conclusioni errate, mentre la determinazione delle correlazioni che tra i sintomi intercorrono è impraticabile senza appropriate rilevazioni statistiche.

Le sistematiche rilevazioni statistiche che qui si auspicano, oltre che permettere di determinare la frequenza con cui le varie costellazioni sintomatiche si presentano nelle diverse forme morbose, potranno servire a stabilire, tra i singoli sintomi, una gerarchia d'importanza e a misurarne le mutue correlazioni, nozioni che, per la scienza come per la pratica, possono riuscire di grande utilità.

Ai fini della diagnosi è però indispensabile considerare le costellazioni sintomatiche complete. Si intende con ciò che tutti i possibili sintomi devono essere presi in esame, in modo che, se la diagnosi si basa su n sintomi sopra $p = n + m$ sintomi possibili, ciò avviene perchè si sono presentati soltanto detti n sintomi, mentre gli m restanti non si sono presentati, pur essendo stati presi in esame. Ciò è essenziale perchè il non presentarsi di un sintomo può costituire un sintomo negativo di valore deci-

sivo. Una costellazione sintomatica che, su possibili sintomi, presenti p - 1 sintomi positivi, ma in cui un sintomo, un solo sintomo, manchi, può essere negativa, escludere, cioè, l'esistenza della malattia in questione, se il mancante è un sintomo che costantemente accompagna la malattia (*sintomo patognomonico negativo*), mentre una costellazione che presenta un solo sintomo positivo può conciliarsi con l'esistenza della malattia se nessuno dei mancanti è sintomo costante della malattia.

È curioso come il Murri, che alla statistica riconosceva la competenza di dare, nel giudizio sull'efficienza dei trattamenti clinici, la sanzione suprema [1], diffidasse invece delle applicazioni del metodo statistico alla diagnostica e ne diffidasse perchè — come egli scriveva: « il valore intrinseco di un fenomeno è poca cosa: invece grande è l'importanza dell'associazione sua con diversi altri fatti ». E soggiungeva: « Nella grande varietà con cui tali fatti si presentano per numero, per intensità, per associazione, per successione, sta l'origine delle difficoltà che incontransi nella pratica » [2]. Se non che la sola conclusione autorizzata da tali considerazioni sarebbe stata che, per superare siffatte difficoltà, occorreva non limitare le rilevazioni statistiche ai singoli sintomi, ma estenderle alle loro successioni ed associazioni, vale a dire alle costellazioni sintomatiche complete.

Vero è che, data la grande varietà delle costellazioni, anche quando queste si vogliano ricondurre con appropriate classificazioni ai tipi fondamentali, grandissimo dovrà essere il numero dei casi raccolti per ricavarne risultati attendibili [3], ma, dovesse anche la raccolta dei dati durare, pur nelle maggiori città, per molti anni, dovesse anche essa giovare solo ai nostri posteri, non perciò sarebbe da rinunciarvi. Lo scienziato deve ormai persuadersi che le più grandi questioni della scienza, e in particolare della biologia, non possono risolversi che con la cooperazione di

---

[1] *Il pensiero scientifico e didattico della Clinica Medica Bolognese.* Due conferenze lette il 14 e il 17 gennaio 1905. Bologna, Foschi, 1905, p. 74.

[2] *Pensieri e precetti*, op. cit. pag. 207.

[3] E ciò tanto più che la probabilità che una data forma morbosa presenti certi sintomi può essere diversa per le varie regioni e per le varie età, le quali devono quindi essere trattate separatamente. Occorre pure fare distinzione tra i sintomi, che possono dirsi *reali*, e i sintomi, in un certo senso *fittizi*, dovuti a forme isteriche.

molti ricercatori, a volte di parecchie generazioni, e deve trovare, nel suo spirito di solidarietà, stimolo sufficiente per tenersi pago di portare la sua modesta pietra ad un edificio di cui forse non potrà mai vedere il coronamento.

Ancora più estesa dovrà essere in molti casi la raccolta dei dati al fine di utilizzare quelle che io chiamerei le *costellazioni sintomatiche composite*, nelle quali, ai sintomi spontaneamente offerti all'osservazione del medico dalle forme morbose, si aggiungono quelli da esso provocati mediante analisi chimiche o con la somministrazione di medicine o con altro trattamento clinico. Ed è in tali casi che il ragionamento istintivo cade facilmente in difetto, evidentemente perchè la nostra facoltà di previsione è sorta attraverso ad una selezione operatasi in condizioni più semplici, nelle quali alla presentazione dei fenomeni doveva seguire pronta l'azione, senza poter attendere il presentarsi di fenomeni complementari. Doveva tornare vantaggioso in tali casi alla specie di utilizzare subito gli elementi di giudizio raccolti, salvo correggerne successivamente le conclusioni raggiunte, senza dover rifare tutto il processo logico. Ora, per giungere a una determinazione corretta della probabilità delle forme morbose, è invece necessario, nelle costellazioni sintomatiche composite, di ritornare da capo, reimpostando il problema in base a tutti i sintomi spontanei e provocati, divenuti in corso di tempo disponibili.

Se, per esempio, in base ai sintomi spontanei la probabilità di una forma morbosa *a* fosse risultata del 60% e quella di una forma morbosa *b* del 30% e, in base alle analisi successivamente compiute, la probabilità di *a* risultasse invece del 20% e quella di *b* dell'80% non sarebbe autorizzato, ai fini di ottenere la probabilità definitiva per la malattia *a*, di fare la media tra 60 e 20 e, per la malattia *b*, la corrispondente media fra 30 e 80. Sarebbe necessario reimpostare il problema e rifare i calcoli, determinando la probabilità che si sia in presenza della malattia *a*, o invece della malattia *b*, quando si presentano congiuntamente i sintomi spontanei e i sintomi provocati che sono stati riscontrati. Data la correlazione, che può esistere o non esistere e, esistendo, assumere diverso grado, tra i sintomi spontanei che si sono dapprima osservati e i sintomi provocati che sono successivamente intervenuti, non è detto affatto che i due procedimenti debbano portare agli stessi risultati.

Il titolo della mia conferenza non contemplava solo le applicazioni della statistica alle diagnosi, ma anche alle prognosi. Le stesse schede raccolte in occasione delle diagnosi potranno anche servire per la determinazione delle probabilità delle prognosi quando su di esse venga poi inscritto lo svolgimento e l'esito della forma morbosa. Qui l'applicazione risulta molto più facile se la prognosi si limita all'esito fausto od infausto del processo morboso, ma, se invece l'esito fausto voglia distinguersi in sotto categorie a seconda della durata della malattia e a seconda che il paziente ne sorta senza menomazione o con qualche menomazione della sua resistenza vitale e, nel secondo caso, a seconda del grado e della durata della menomazione, e se, analogamente, lo esito infausto si distingua in sottocategorie secondo la durata del processo morboso, le alternative da prendere in considerazione (corrispondenti alle « cause » nella determinazione della probabilità a posteriori) possono divenire non meno numerose delle possibili forme morbose da prendere in considerazione nella determinazione della diagnosi, con l'aggravante che la distinzione fra alcune alternative risulta più o meno arbitraria, e l'applicazione della statistica alle prognosi può divenire in pratica altrettanto, se non più, delicata dell'applicazione alle diagnosi.

Sulla via da percorrere il primo passo sarà compiuto, arra di felice proseguimento, se verrà accettata la proposta, con cui chiudo il mio dire, che una ristretta commissione venga costituita, composta di clinici e di statistici, al fine di preparare una scheda clinica adatta, o di modificare convenientemente quella già in uso, così da raccogliere le indicazioni occorrenti per ricavarne i dati statistici necessari ad una rigorosa determinazione quantitativa delle diagnosi e delle prognosi.

CARLO BENEDETTI

Di un confronto fra poteri di acquisto di salari e considerazioni metodologiche relative agli indici circolari e non circolari.

Lo scopo del presente lavoro è duplice: economico e metodologico.

Il primo consiste nel confrontare fra loro i poteri di acquisto dei salari nel 1938 di 7 città: *Bolzano, Lecce, Salerno, Messina, Campobasso, Milano, Sassari* in modo da poter disporre le dette città in una graduatoria conclusiva dove esse siano disposte secondo il livello decrescente del rispettivo potere di acquisto (vedi prospetto D).

Il secondo, dopo aver esposto lo schema di calcolo adoperato ed aver accennato al cosidetto *indice dei tempi* (indice di un certo interesse), consiste nel far notare che per giungere a risultati non contraddittori è necessario adoperare nel calcolo degli indici del potere di acquisto formule soddisfacenti alla cosidetta proprietà *circolare*; tali sono le 4 formule del prof. Gini [1] che abbiamo adoperato nell'indagine numerica; inoltre, sempre a questo proposito, si dimostra un *teorema generale* dove è provato rigorosamente che tutti gli indici non circolari di una certa forma (praticamente tutti gli indici non circolari conosciuti) fra cui la cosidetta « formula ideale », danno luogo a contraddizioni.

Proprio nella nostra indagine fra le 7 città vedremo che la « formula ideale » per le contraddizioni cui dà luogo, non consente di effettuare la graduatoria dei poteri di acquisto.

---

(1) C. Gini - *On the Circular test of index numbers* « Metron » 1931-32 Vol. IX.

Infine si danno delle condizioni affinchè un indice soddisfi alla proprietà circolare e confrontiamo lo schema che ne deriva con quello di Ragnar Frisch [1] diversamente impostato e le cui formule soddisfano ciascuna simultaneamente un numero di certe proprietà minore di quello soddisfatto simultaneamente da ciascuna delle formule del Gini e di altre del nostro schema.

Risultati di carattere economico

1. — I poteri di acquisto da confrontare fra loro si riferiscono ai salari di 2 categorie di lavoratori: *Metalmeccanici* ed *Edili*; ogni categoria è suddivisa in 4 classi:

I operaio specializzato
II operaio qualificato
III manovale specializzato
IV manovale comune

Le unità di merce composita o « cestini » rispetto a cui si misura il potere di acquisto dei vari salari, sono costituite da 7 razioni alimentari medie giornaliere espresse in grammi per unità-uomo [2]. Tali razioni sono state rilevate nell'aprile-maggio 1929 per le città: Bolzano, Lecce, Salerno, Sassari, Messina, Campobasso e Milano a seguito dell'inchiesta alimentare eseguita per conto della *Commissione per lo studio dei problemi dell'alimentazione* del Consiglio Nazionale delle Ricerche; i risultati di tale inchiesta sono esposti in 5 volumi (a cura dello stesso Consiglio Nazionale delle Ricerche) dal titolo: Quaderni della Nutrizione.

---

[1] Ragnar Frisch - *Necessary and sufficient conditions regarding the form of an index number which shall meet certain of Fisher's tests* - in « Journal of the American Statistical Association » del Dicembre 1930.

[2] Il numero di unità-uomo attribuibile a ciascuna famiglia è ottenuto applicando al numero effettivo dei componenti la cosidetta scala del Lusk che fissa i seguenti coefficienti:

da 0 a 6 anni.	0,50
da 6 a 10 anni.	0,70
da 10 a 14 anni.	0 83
da 14 a ω anni.	0,83 per la donna 1 per l'uomo

Così per avere la razione media giornaliera il totale consumo delle famiglie rilevate per ogni categoria interessata viene diviso per le complessive unità-uomo a cui sono ridotti gli effettivi componenti le famiglie stesse. (Vedi: Quaderni della Nutrizione).

Nei calcoli dei vari indici abbiamo adottato le razioni medie giornaliere della voce 7 delle tabelle dei Quaderni della Nutrizione e cioè quella degli «operai delle industrie, arti e mestieri» in modo da centrare al massimo le due categorie (Metalmeccanici ed Edili) scelte per i salari.

Di più, è da notare che in tutti i nostri calcoli insieme alle razioni dell'aprile-maggio 1929, si sono adottati i prezzi dei rispettivi generi alimentari componenti dell'aprile 1938 ed i salari anch'essi del 1938 (supporremo cioè invariate nel 1938 le razioni del 1929); ciò per le seguenti ragioni:

a) Non abbiamo potuto avere per tutte le 7 città i prezzi locali al minuto del 1929 e di anni ad esso vicini nè le razioni del 1938.

b) Non abbiamo applicato alle razioni di 23 anni orsono i prezzi e salari di quest'anno o di qualche anno fa a causa di presumibili forti variazioni nella composizione per quantità e qualità delle razioni stesse.

I prezzi locali al minuto per il 1938 sono stati tratti per alcuni generi dai bollettini mensili dei prezzi, per altri generi, per i quali tali bollettini non contemplavano alcun prezzo, ci siamo rivolti alle rispettive Camere di Commercio oppure agli Uffici Statistici dei vari Comuni i quali, come è noto, compiono tali rilevazioni seguendo svariatissimi criteri.

Per i salari abbiamo preso le retribuzioni minime contrattuali rilevate sui contratti collettivi di lavoro a cura della Confindustria per i Metalmeccanici ed a cura della Associazione Nazionale Costruttori Edili per gli Edili.

2. — Come è già stato detto per ogni città abbiamo 2 categorie di lavoratori; Metalmeccanici ed Edili ed ogni categoria è suddivisa in 4 classi: I operai specializzati; II operai qualificati; III manovali specializzati; IV manovali comuni.

Nel calcolo dei vari indici (dei salari, potere di acquisto e tempi) alla classe i-esima della categoria j-esima di una città si contrappone sempre la stessa classe i-esima della stessa categoria j-esima di ogni altra città.

Così, mediante le formule del Gini [1] troviamo prima gli indici dei prezzi di una città rispetto a ciascuna altra (questi dei

---

[1] Lo schema di calcolo e formule relative verranno esposti nella seconda parte fra le considerazioni metodologiche.

prezzi sono indici ovviamente indipendenti dalle classi e categorie).

Poi troviamo gli indici dei salari di ogni città rispetto a ciascun'altra (mediante semplici rapporti).

Infine troviamo gli indici del potere di acquisto di ogni città rispetto a ciascun'altra dividendo per ogni coppia di città (coppia di classi, di categorie, di città) il rispettivo indice dei salari per il corrispondente indice dei prezzi.

Nelle tabelle 1ª 2ª 3ª 4ª e 5ª che alleghiamo sono indicati i valori degli indici dei prezzi, del potere di acquisto e dei tempi (l'indice dei tempi è l'inverso di quello del potere di acquisto e lo illustreremo più avanti) ottenuti nelle prime quattro tabelle con le formule del Gini e nella 5ª con quella « ideale ».

Nella tabella 5ª bis sono gli indici dei salari ottenuti come abbiamo già accennato mediante semplici rapporti.

Tutte queste tabelle sono a doppia entrata in modo da far corrispondere ad ogni città tutte le altre compresa se stessa.

Le città sulla testata orizzontale stanno a numeratore cioè gli indici che si trovano nella colonna intestata ad una di esse ci dicono a seconda se maggiori o minori di 1 se la città in questione rispettivamente supera o no (in prezzi, in salari, in potere di acquisto ecc...) le altre che incontra lungo la verticale.

Le prime 4 tabelle : 1ª, 2ª, 3ª e 4ª contengono l'applicazione agli stessi dati delle 4 formule del Gini che sono sostanzialmente equivalenti (hanno in comune la proprietà circolare ed altre proprietà come vedremo in seguito) quindi per ogni indice particolare abbiamo 4 valori equivalenti che sono stati tabulati distintamente per sottolineare la notevole concordanza esistente entro le quaterne stesse, significando ciò, in ultima analisi, conferire sufficiente grado di attendibilità ai risultati comuni raggiunti con le quattro formule.

3. — Per avere un'idea più chiara possibile della posizione di ogni città rispetto a ciascun'altra per il potere di acquisto e prezzi abbiamo eseguito, per ognuna delle 4 formule adoperate le graduatorie dei risultati ottenuti (vedi prospetto A).

Per ogni tabella ottenuta con formula del Gini è possibile infatti per ogni indice (dei prezzi, potere di acquisto ecc.) fare la graduatoria delle città disponendole secondo l'altezza crescente o decrescente del livello dei prezzi, potere di acquisto, ecc..

Volendo ad esempio graduare le 7 città secondo il livello dei prezzi basta per ogni colonna di una delle 4 tabelle allegate contare il numero delle volte in cui l'indice dei prezzi (della città intestataria della colonna rispetto a ciascun'altra) risulta inferiore ad 1, oppure il numero delle volte in cui risulta superiore.

Detto n_1 il primo numero ed n_2 il secondo, volendo graduare le suddette città secondo ad esempio l'altezza decrescente del livello dei prezzi, dovremo attribuire all'ordine in graduatoria della città intestataria della colonna il numero $7 - n_2$ oppure $n_1 + 1$ poichè sarà sempre $n_1 + n_2 = 6$ (perchè trascuriamo la posizione di ogni città rispetto a se stessa).

Questa operazione di graduazione non è sempre possibile se gli indici sono ottenuti con la formula « ideale » ed altre formule non circolari ; dimostreremo rigorosamente ciò nelle considerazioni di carattere metodologico.

Ora, allo scopo di appurare la concordanza tra i risultati delle quattro formule del Gini si sono calcolati gli indici semplici di cograduazione [1] (vedi prospetto *B*) tra ogni graduatoria dei risultati ottenuti con una data formula e la corrispondente graduatoria di risultati ottenuti con ognuna delle 3 rimanenti formule ; dalla forte cograduazione che si nota nel prospetto è naturale concludere che i risultati delle 4 formule si confermano a vicenda.

Nel prospetto *B* non abbiamo calcolato gli indici di cograduazione dei salari perchè questi come è noto non variano da formula a formula.

A questo punto, data la notevole concordanza dei corrispondenti risultati delle 4 formule, si è ritenuta legittima la sintesi delle 4 posizioni di ogni singola classe di lavoratori in una posizione media unica.

A tale scopo abbiamo compilato il prospetto *C* nel quale per ogni indice (dei prezzi, salari, poteri di acquisto e tempi) in ogni classe di ogni città è stata data una posizione di graduatoria media aritmetica tra le 4 posizioni che la stessa classe aveva secondo le 4 formule adoperate.

---

[1] Che permettono di misurare il grado di concordanza tra i posti delle graduatorie che le 7 città occupano nelle prime 4 tabelle.

Vedi ad esempio: C. Gini - *Corso di Statistica* 1952-53; Roma, Veschi.

Ora, siccome lo scopo del presente lavoro è di confrontare fra loro i poteri di acquisto (che coinvolgono salari e prezzi) delle varie città, abbiamo estratto dal prospetto *C* un altro prospetto *D* dove si sono riportate le 8 graduatorie (una per classe di lavoratori) del potere di acquisto nelle 7 città.

L'ordine dato alle città in questo nuovo prospetto è stato ottenuto nel modo seguente:

Per ogni gruppo di 8 numeri naturali a fianco di ogni città è stata fatta la media aritmetica ed ordinato le città secondo i valori crescenti di tali medie; ecco il prospetto *D*:

	MI	*MII*	*MIII*	*MIV*	*CI*	*CII*	*CIII*	*CIV*
1 - Milano	1	2	1	1	1	1	1	1
2 - Bolzano	2	1	2	2	2,5	2	2,5	2
3 - Sassari	5	3	7	—	2,5	4	2,5	3,5
4 - Salerno.	4	6	5	5	5	5	4	3,5
5 - Campobasso. . .	3	5	3	4	7	6	7	6
6 - Messina.	7	4	4	3	6	7	6	5
7 - Lecce.	6	7	6	6	4	3	5	7

dove: *M* = Metalmeccanici; *C* = Edili;

I = operaio specializzato; II = operaio qualificato;

III = manovale specializzato; IV = manovale comune.

A Sassari manca la posizione in graduatoria di *M* IV perchè per tale classe non abbiamo il salario orario e quindi nemmeno gli indici del potere di acquisto di essa rispetto alla stessa classe *M* IV delle altre città.

Per ordinare al terzo posto Sassari abbiamo diviso per 7 la somma dei posti di ordine delle sue classi, e ciò per ovvie ragioni di comparabilità.

Precisato questo vediamo *che Milano ha in media il potere di acquisto più alto seguito da Bolzano, Sassari, Salerno, Campobasso, Messina e Lecce.*

Il fatto che Sassari si sia piazzato al 3 posto dipende (come si può vedere subito nel prospetto *A*) dai suoi prezzi molto bassi, ultimi in graduatoria, ai quali si contrappongono i suoi salari non sempre ultimi in graduatoria.

Infine per completare l'esame del prospetto *D* abbiamo calcolato gli indici di cograduazione tra le 8 graduatorie del prospetto stesso allo scopo di conoscere le classi fra le quali esiste la massima cograduazione nelle diverse città.

Tali indici figurano nel prospetto *E* seguente ;

			II				*I*	
	MI	MII	MIII	MIV	CI	CII	CIII	CIV
MI	1, —	0,333	0,625	0,556	0,375	0,500	0,500	0,583
MII. . . .	0,333	1, —	0,458	0,889	0,542	0,375	0,542	0,625
MIII . . .	0,625	0,458	1, —	0,778	0,250	0,333	0,250	0,418
MIV . . .	0,556	0,889	0,778	1, —	0,444	0,444	0,444	0,667
CI	0,375	0,542	0,250	0,444	1, —	0,792	0,917	0,625
CII	0,500	0,375	0,333	0,444	0,792	1, —	0,625	0,542
CIII . . .	0,500	0,542	0,250	0,444	0,917	0,625	1, —	0,750
CIV. . . .	0,583	0,625	0,418	0,667	0,625	0,542	0,750	1, —
			III				*IV*	

(Il quadrante *I* è identico al quadrante *III* per ovvie ragioni)

e mostrano nei 4 quadranti in cui sono stati suddivisi che i massimi sono raggiunti quasi uniformemente nel *IV* quadrante e cioè fra le 4 classi degli Edili ; i minimi sono toccati nel *I* e *III* quadrante (identici) dove si trovano confrontate le classi degli Edili con quelle dei Metalmeccanici .

Considerazioni di carattere metodologico.

1. *Schema di calcolo adoperato.* — Anzitutto si sono calcolati gli indici dei prezzi di ogni città rispetto a ciascun'altra mediante le 4 formule del Gini e che compaiono per la prima volta nell'articolo già citato : « On the circular test of index numbers » Metron vol. IX 1931-32.

Cominciando dalle prime due che sono strutturalmente analoghe e indicando con ${}_{a,b}I_p$ l'indice dei prezzi della città a rispetto alla città b abbiamo:

$$\text{(I)}\quad {}_{a,b}I_p = \frac{\sum_{l=1}^{s} P_{al}\,\frac{1}{n}\sum_{i=1}^{n} Q_{il}}{\sum_{l=1}^{s} P_{bl}\,\frac{1}{n}\sum_{i=1}^{n} Q_{il}}\;;\qquad \text{(II)}\quad {}_{a,b}I_p = \frac{\sum_{l=1}^{s} P_{al}\,\frac{\sum_{i=1}^{n} P_{il}\,Q_{il}}{\sum_{i=1}^{n} P_{il}}}{\sum_{l=1}^{s} P_{bl}\,\frac{\sum_{i=1}^{n} P_{il}\,Q_{il}}{\sum_{i=1}^{n} P_{il}}}$$

dove P_{al}, P_{bl}, P_{il}, Q_{il} sono tre prezzi ed una quantità del genere alimentare l rispettivamente nelle città a, b, i;
n è il numero delle città (nel nostro caso n $= 7$);
s è il numero dei generi alimentari componenti la razione (nel nostro caso $s = 17$).

Nelle due formule precedenti per ogni genere alimentare prendiamo la media aritmetica rispettivamente semplice e ponderata delle 7 quantità (tante quante le città) che di tale genere figurano nelle 7 razioni, in modo che la quantità di uno stesso genere alimentare è costante per tutte le coppie di città confrontate.

Le altre due formule sono:

$$\text{(III)}\qquad {}_{a,b}I_p = \sqrt[n]{\frac{\sum_{l=1}^{s} P_{al}\,Q_{al}}{\sum_{l=1}^{s} P_{bl}\,Q_{al}}\;\frac{\sum_{l=1}^{s} P_{al}\,Q_{bl}}{\sum_{l=1}^{s} P_{bl}\,Q_{bl}}\;\frac{\sum_{l=1}^{s} P_{al}\,Q_{cl}}{\sum_{l=1}^{s} P_{bl}\,Q_{cl}}\cdots\cdots\frac{\sum_{l=1}^{s} P_{al}\,Q_{zl}}{\sum_{l=1}^{s} P_{bl}\,Q_{zl}}}$$

che é una notevole estensione della « formula ideale »; e

$$\text{(IV)}\qquad {}_{a,b}I_p = \sqrt[n]{\frac{a}{b}\frac{a}{a}\;\frac{a}{b}\frac{b}{b}\;\frac{a}{b}\frac{c}{c}\cdots\cdots\frac{a}{b}\frac{z}{z}}$$

per i P ed i Q, per s, l, n valgono ovviamente gli stessi significati che della (I) e (II); di più nella (IV) è

$$\frac{a}{a}\,a = 1\;;\; \frac{a}{b}\,b = \sqrt{\frac{\sum_{l=1}^{s} P_{al}\,Q_{al}}{\sum_{l=1}^{s} P_{bl}\,Q_{al}}\;\frac{\sum_{l=1}^{s} P_{al}\,Q_{bl}}{\sum_{l=1}^{s} P_{bl}\,Q_{bl}}}\;;\ \text{ecc.}$$

cioè $\underline{a\,a}$, $\underline{a\,b}$, ... $\underline{a\,z}$ sono indici dei prezzi ottenuti mediante la « formula ideale ».

Per queste 4 formule, come è facile vedere, vale la proprietà circolare e cioè: essendo a, b, x tre tempi o luoghi presi a caso tra n, distinti o no è sempre:

$$(\alpha) \qquad \frac{{}_{a,x}I_p}{{}_{b,x}I_p} = {}_{a,b}I_p \; ; \quad \text{ora, ponendo } x = b \text{ avremo:}$$

$$(\alpha') \qquad \frac{{}_{a,b}I_p}{{}_{b,b}I_p} = {}_{a,b}I_p \quad \text{che implica } {}_{b,b}I_p = 1 \; ;$$

in (α) ponendo ora $x = a$ avremo per la (α')

$$(\alpha'') \qquad \frac{{}_{a,a}I_p}{{}_{b,a}I_p} = {}_{a,b}I_p = \frac{1}{{}_{b,a}I_p} \; .$$

Quindi la proprietà circolare (espressa in (α)) implica necessariamente la proprietà di identità (*identity test*) espressa in (α') e la proprietà di reversibilità dei tempi e luoghi (*time* oppure *place reversal test*) espressa in (α''). Non è vero il contrario come vedremo fra poco.

La (α) possiamo evidentemente scriverla così:

$$ {}_{a,x}I_p \cdot {}_{x,b}I_p \cdot {}_{b,a}I_p = 1$$

dove a, b, x é una terna qualunque tra le $\binom{n+2}{3}$ (combinazioni con ripetizione)

L'indice del potere acquisto ${}_{a,b}I_A$ della città a rispetto alla città b è ottenuto per ogni categoria e classe di lavoratori dividendo per il rispettivo indice dei prezzi ${}_{a,b}I_p$ il rapporto $\frac{S_a}{S_b}$ dove S_a è il Salario orario in a ed S_b quello in b.

Allora sarà, per le tre città generiche a, b, c

$${}_{a,b}I_A = \frac{\frac{S_a}{S_b}}{{}_{a,b}I_p} \; ; \; {}_{b,c}I_A = \frac{\frac{S_b}{S_c}}{{}_{b,c}I_p} \; ; \; {}_{a,c}I_A = \frac{\frac{S_a}{S_c}}{{}_{a,c}I_p}$$

da cui :

$$_{a,c}I_A = {}_{a,b}I_A \cdot {}_{b,c}I_A ; \qquad (\beta)$$

cioè anche per gli indici del potere di acquisto si verifica la proprietà circolare in quanto si tratta di quozienti di indici in cui tale proprietà è verificata ; infatti per l'indice dei salari è :

$$\frac{S_a}{S_c} = \frac{S_a}{S_b}\frac{S_b}{S_c} = \frac{\dfrac{S_a}{S_b}}{\dfrac{S_c}{S_b}}$$

Ritorneremo sulla proprietà circolare più diffusamente in seguito.

Brevemente accenneremo ora ad un indice (anch'esso considerato nelle tabelle e prospetti allegati) che ha un significato di un certo interesse : *l'indice dei tempi*.

Se nelle 4 formule adoperate, in ciascuna di esse dividiamo ogni prodotto PQ in senso lato (cioè ogni prodotto di un prezzo per una quantità effettiva o media) per il salario orario della città cui appartiene P otteniamo il tempo di lavoro necessario per conseguire nella città in questione la somma PQ.

Mediante questi cambiamenti dell'unità di misura negli importi suddetti l'indice dei prezzi ${}_{a,b}I_p$ si trasforma nell'indice dei tempi ${}_{a,b}I_t$ che ci dice in media se il tempo impiegato a guadagnare la somma necessaria per acquistare la razione giornaliera in a è superiore od inferiore, ed in quale misura, a quello occorrente ad acquistare la razione in b.

Infatti nella (I) avremo :

$$_{a,b}I_t = \frac{S_b}{S_a}\frac{\sum_{l=1}^{s} P_{al}\frac{1}{n}\sum_{i=1}^{n} Q_{il}}{\sum_{l=1}^{s} P_{bl}\frac{1}{n}\sum_{i=1}^{n} Q_{il}} = \frac{S_b}{S_a}\frac{\sum_{l=1}^{s} P_{al}\overline{Q}_l}{\sum_{l=1}^{s} P_{bl}\overline{Q}_l} = \frac{\frac{1}{S_a}\sum_{l=1}^{s} P_{al}\overline{Q}_l}{\frac{1}{S_b}\sum_{l=1}^{s} P_{bl}\overline{Q}_l} = \frac{\sum_{l=1}^{s} t_{al}}{\sum_{l=1}^{s} t_{bl}} =$$

$$= \frac{\text{totale tempi in } a}{\text{totale tempi in } b} \left(\text{dove } \overline{Q}_l = \frac{1}{n}\sum_{i=1}^{n} Q_{il} ;\ t_{a_l} = \frac{P_{al}\overline{Q}_l}{S_a} ;\ ecc. \right)$$

lo stesso per la (II) quando si tenga presente che in questa sarà

$$\overline{Q}_l = \frac{\Sigma P_{il} Q_{il}}{\Sigma P_{il}}$$

Nella (III) avremo :

$$ {}_{a,b}I_t = \frac{S_b}{S_a} \sqrt[n]{\prod_{i=1}^{n} \frac{\sum_{l=1}^{s} P_{al} Q_{il}}{\sum_{l=1}^{s} P_{bl} Q_{il}}} = \sqrt[n]{\prod_{i=1}^{n} \frac{\frac{1}{S_a} \sum_{l=1}^{s} P_{al} Q_{il}}{\frac{1}{S_b} \sum_{l=1}^{s} P_{bl} Q_{il}}} = \sqrt[n]{\prod_{i=1}^{n} \frac{\sum_{l=1}^{s} t_{al,i}}{\sum_{l=1}^{s} t_{bl,i}}} $$

Che è la media geometrica dei rapporti fra il tempo occorrente per guadagnare la razione in *a* ed il tempo per guadagnare la razione in *b* facendo variare i cestini nei 7 modi possibili.

Così per la (IV) se cominciamo ad esprimere nei rispettivi tempi di lavoro i prodotti PQ che figurano negli indici elementari $\underline{aa}$, $\underline{ab}$, $\underline{ac}$, ... ecc. ottenuti con la « formula ideale » arriviamo alla stessa interpretazione delle tre formule precedenti.

Ora si vede subito che l'inverso di ${}_{a,b}I_t$ testé illustrato è l'indice del potere di acquisto ${}_{a,b}I_A$ definito direttamente più indietro.

Nelle tabelle 1ª, 2ª, 3ª, 4ª, 5ª, 5ªbis e nei prospetti allegati abbiamo tabulato i valori e le graduatorie di I_t, I_p, I_A ; $\frac{S_a}{S_b}$.

2. *Carattere di necessità della proprietà circolare.* La proprietà circolare più sopra definita, non è soltanto una elegante proprietà formale ma risponde ad una elementare ed insopprimibile esigenza logica.

Infatti basta osservare che se per un qualunque indice abbiamo ad esempio ${}_{a,b}I_p = 2$; ${}_{b,c}I_p = 2$ dovrà essere necessariamente ${}_{a,c}I_p = 2 \times 2 = 4$ e non altrimenti e cioè ${}_{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p$; così anche per le altre due proprietà connesse alla circolare infatti se ad esenpio è ${}_{a,b}I_p = 2$ sarà necessariamente ${}_{b,a}I_p = \frac{1}{2} = \frac{1}{{}_{a,b}I_p}$ (reversibilità) ; e così dovrà anche essere ${}_{a,a}I_p = {}_{b,b}I_p = {}_{c,c}I_p = \ldots = 1$. (propr. di identità)

Quando un indice non soddisfa alla proprietà circolare non è quindi coerente con la logica elementare.

Ora è facile convincersi, come è già stato visto, che la proprietà circolare implica necessariamente e congiuntamente le proprietà di reversibilità (tempi e luoghi) e di identità ma che non è vero il contrario ; cioè esistono indici che pur soddisfacendo a queste due ultime proprietà non soddisfano quella circolare ; tale è il caso della « formula ideale ». Faremo vedere tra poco le contraddizioni cui possono condurre tali indici non circolari.

Nella « formula ideale » infatti si ha :

$$_{a,b}I_p \cdot {}_{b,c}I_p = {}_{a,c}I_p \sqrt{\frac{\Sigma P_a Q_b \, \Sigma P_b Q_c \, \Sigma P_c Q_a}{\Sigma P_c Q_b \, \Sigma P_a Q_c \, \Sigma P_b Q_a}}$$

oppure che è lo stesso

$$_{a,c}I_p \cdot \sqrt{\frac{\Sigma P_a Q_b}{\Sigma P_c Q_b}} = {}_{a,b}I_p \cdot \sqrt{\frac{\Sigma P_a Q_c}{\Sigma P_b Q_c}} \cdot {}_{b,c}I_p \cdot \sqrt{\frac{\Sigma P_b Q_a}{\Sigma P_c Q_a}}$$

in cui vediamo che si può arrivare alla proprietà circolare includendo in ogni indice il « cestino » mancante e ciò conduce spontaneamente alla (III) del Gini mediante il cambiamento della radice quadrata in radice cubica per dare la giusta dimensione agli indici. Siccome nel caso di n città la proprietà circolare si deve verificare per tutte le terne $\left[\text{in numero di } \binom{n+2}{3}\right]$ è immediata la generalizzazione includendo in ogni indice gli n cestini ed estrarre la radice n-esima.

Ora, dalla relazione (β) che riscriveremo :

$$_{a,c}I_A = {}_{a,b}I_A \cdot {}_{b,c}I_A$$

deriva ovviamente che se

$$\left.\begin{aligned} {}_{a,b}I_A > 1 \text{ e } {}_{b,c}I_A > 1 \quad &\text{sarà anche} \quad {}_{a,c}I_A > 1 \\ {}_{a,b}I_A < 1 \text{ e } {}_{b,c}I_A < 1 \quad &\text{sarà anche} \quad {}_{a,c}I_A < 1 \end{aligned}\right\} \ (\beta)'$$

Adoperando la « formula ideale » ed una infinità di formule *non circolari* di un certo tipo (come vedremo fra poco) non si soddisfa in generale alla (β) e non sempre alle (β)'.

Come si nota subito il non soddisfare alla (β)' è molto più grave ai fini delle graduatorie dei risultati (come noi abbiamo fatto) che non soddisfare alla (β)

Ora, nella tabella 5ª (dati secondo la « formula ideale ») nella classe I dei metalmeccanici fra le città Lecce $= (L)$, Salerno $= (S)$ e Messina $= (M)$ abbiamo :

$$_{(S),(L)}I_p = 1{,}051 \ldots {}_{(L),(M)}I_p = 0{,}996 \ldots {}_{(S),(M)}I_p = 0{,}942$$

da cui

$$1{,}051.. \times 0{,}996.. = 1{,}047 \quad \text{anzichè} \quad 0{,}942...$$

Questo fatto (cioè la discordanza di 0,942 da 1,047) si ripercuote e viene aggravato negli indici dei poteri di acquisto dove dividiamo gli indici dei salari (fra cui vale la proprietà circolare) per gli I_p (fra cui non vale) si ha così:

$$\frac{S_{(S)}}{S_{(L)}} = 1.044... ; \frac{S_{(L)}}{S_{(M)}} = 0{,}937... ; \frac{S_{(S)}}{S_{(M)}} = 0{,}978....$$

da cui:

$${}_{(S),(L)}I_A = \frac{1{,}044..}{1{,}051..} = 0{,}994.. ; {}_{(L),(M)}I_A = \frac{0{,}937...}{0.996...} = 0.940 ;$$

$${}_{(S),(M)}I_A = \frac{0{,}978..}{0{,}942..} = 1{,}038..$$

che sono tre risultati contraddittori perchè essendo il potere di acquisto di Salerno *inferiore* a quello di Lecce e quello di Lecce *inferiore* a quello di Messina, si ha che quello di Salerno è assurdamente *superiore* a quello di Messina il che in simboli si traduce nella contraddizione:

$$A < B, \quad B < C, \rightarrow A > C$$

Il caso numerico concreto di contraddizione cui ha dato luogo la « formula ideale » è il particolare verificarsi di una possibilità la cui esistenza è provata rigorosamente dal teorema seguente dove si dimostra sostanzialmente che tutti gli indici non circolari possono condurre a contraddizioni del tipo:

$$A > B, B > C, \rightarrow A < C$$

oppure

$$A < B, \quad B < C, \rightarrow A > C$$

Teorema — Date 6 successioni di numeri reali positivi:

$$p_{a1},\ p_{a2},\ \ldots\ldots\ldots\ldots\ p_{as}$$
$$p_{b1},\ p_{b2},\ \ldots\ldots\ldots\ldots\ p_{bs}$$
$$p_{c1},\ p_{c2},\ \ldots\ldots\ldots\ldots\ p_{cs}$$

$$q_{a1},\ q_{a2},\ \ldots\ldots\ldots\ldots\ q_{as}$$
$$q_{b1},\ q_{b2}\ \ldots\ldots\ldots\ldots\ q_{bs}$$
$$q_{c1},\ q_{c2},\ \ldots\ldots\ldots\ldots\ q_{cs}$$

costituenti i prezzi: (p_{ai}), (p_{bi}), (p_{ci}) e le quantità (q_{ai}), (q_{bi}), (q_{ci}) inerenti agli s bene considerati in ciascuno dei tre tempi o luoghi distinti: a, b, c
e dato un indice *non circolare dei prezzi*:

$$_{x,y}I_p = F\,(p_x\,,\ p_y\,;\ q_x\,,\ q_y)$$

contenente solamente prezzi e quantità relativi ai due tempi o luoghi confrontati x e y, e soddisfacente alla condizione:

(1) $$F\,(k\,p_x\,,\ h\,p_y\,;\ q_x\,,\ q_y) = \frac{k}{h}\,F\,(p_x\,,\ p_y\,;\ q_x\,,\ q_y) = \frac{k}{h}\,{}_{x,y}I_p\,;$$

con k ed h costanti reali qualunque:
— possiamo sempre trovare altre due successioni di reali positivi:

$$p'_{a1}\,,\ p'_{a2}\,,\ \ldots\ldots\ldots\ p'_{as}$$

$$p'_{c1}\,,\ p'_{c2}\,,\ \ldots\ldots\ldots\ p'_{cs}$$

tali che si abbiano le contraddizioni:

$$_{a,b}I_{p'} > 1\ ,\ {}_{b,c}I_{p'} > 1 \rightarrow {}_{a,c}I_{p'} < 1$$

oppure:

$$_{a,b}I_{p'} < 1\ ,\ {}_{b,c}I_{p'} < 1 \rightarrow {}_{a,c}I_{p'} > 1$$

dove:

$$_{a,b}I_{p'} = F\,(p'_a\,,\ p_b\,;\ q_a\,,\ q_b)\,;\ {}_{b,c}I_{p'} = F\,(p_b\,,\ p'_c\,;\ q_b\,,\ q_c)\,;$$

$$_{a,c}I_{p'} = F\,(p'_a\,,\ p'_c\,;\ q_a\,,\ q_c)$$

Osservazione: Vediamo che rientrano in questo tipo tutti gli indici non circolari conosciuti tra cui la « formula ideale » e tutte quelle del tipo:

$$\frac{\Sigma P_a M_{q_a, q_b}}{\Sigma P_b M_{q_a, q_b}}$$

dove M_{q_a, q_b} è una media qualunque di q_a e q_b, ecc....

Dimostrazione: Elencando ora, i casi in cui gli indici *non circolari* suddetti *non danno luogo a* contraddizione avremo: (1)

$$1^0 \quad {}_{a,b}I_p > 1 \; ; \; {}_{b,c}I_p > 1 \; ; \; {}_{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p \, (1 - \varepsilon) > 1$$

$$2^0 \quad {}_{a,b}I_p > 1 \; ; \; {}_{b,c}I_p > 1 \; ; \; {}_{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p \, (1 + \varepsilon) > 1$$

$$3^0 \quad {}_{a,b}I_p < 1 \; ; \; {}_{b,c}I_p < 1 \; ; \; {}_{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p \, (1 - \varepsilon) < 1$$

$$4^0 \quad {}_{a,b}I_p < 1 \; ; \; {}_{b,c}I_p < 1 \; ; \; {}_{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p \, (1 + \varepsilon) < 1$$

$$5^0 \quad {}_{a,b}I_p < 1 \; ; \; {}_{b,c}I_p > 1 \; ; \; {}_{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p \, (1 - \varepsilon) \gtrless 1$$

$$6^0 \quad {}_{a,b}I_p < 1 \; ; \; {}_{b,c}I_p > 1 \; ; \; {}_{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p \, (1 + \varepsilon) \gtrless 1$$

con ε compreso fra o ed 1 (estremi esclusi) in $1 - \varepsilon$ ed ε semplicemente maggiore di o in $1 + \varepsilon$

Ora faremo vedere caso per caso che mantenendo le stesse quantità (q_{ai}), (q_{bi}), (q_{ci}) ed i prezzi p_{bi}, possiamo trovare dei reali positivi (p'_{ai}) e (p'_{ci}) da sostituire ai rispettivi (p_{ai}) e (p_{ci}) in modo *da trasformare ognuno dei casi precedenti in una contraddizione.*

1^0 *Caso* ${}_{a,b}I_p > 1 \; ; \; {}_{b,c}I_p > 1 \; ; \; {}_{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p \, (1-\varepsilon) > 1 \; ; (0 < \varepsilon < 1)$

Allora prendendo:

$$p'_{ai} = \frac{p_{ai}}{(1 - \varepsilon_1) \, {}_{a,b}I_p} \; ; \; p'_{ci} = p_{ci} \, (1 - \varepsilon_2) \, {}_{b,c}I_p \; ; \qquad (i = 1, 2, 3, \ldots s)$$

con

$$0 < \varepsilon_1 , \varepsilon_2 < 1$$

(1) Si tralasciano per brevità i casi in cui compaiono anche uguaglianze.

avremo che, tenendo presente la (1)

$$_{a,b}I_{p'} = F(p'_a, p_b; q_a, q_b) = \frac{F(p_a, p_b; q_a, q_b)}{(1-\varepsilon_1)\ _{a,b}I_p} = \frac{_{a,b}I_p}{(1-\varepsilon_1)\ _{a,b}I_p} = \frac{1}{1-\varepsilon_1} > 1$$

analogamente si ha

$$_{b,c}I_{p'} = \frac{_{b,c}I_p}{(1-\varepsilon_2)\ _{b,c}I_p} = \frac{1}{1-\varepsilon_2} > 1\ ;\ _{a,c}I_{p'} = \frac{_{a,c}I_p}{(1-\varepsilon_1)(1-\varepsilon_2)\ _{a,b}I_p \cdot\ _{b,c}I_p}$$

da cui poichè

$$_{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p\,(1-\varepsilon) > 1$$

ricaviamo:

$$_{a,c}I_{p'} = \frac{_{a,b}I_p \cdot {}_{b,c}I_p\,(1-\varepsilon)}{(1-\varepsilon_1)(1-\varepsilon_2)\ _{a,b}I_p \cdot\ _{b,c}I_p} = \frac{1-\varepsilon}{(1-\varepsilon_1)(1-\varepsilon_2)}$$

ora, affinchè sia $_{a,c}I_{p'} < 1$ basta prendere:

$$\varepsilon_1 = \varepsilon_2 = \frac{\varepsilon}{2}$$

per cui:

$$_{a,c}I_p = \frac{1-\varepsilon}{\left(1-\frac{\varepsilon}{2}\right)^2} = \frac{1-\varepsilon}{1-\varepsilon+\frac{\varepsilon^2}{4}} < 1$$

Abbiamo così per certi (p'_{ai}) e (p'_{ci}) e con gli stessi (p_{bi}), (q_{ai}), (q_{bi}), (q_{ci}) la contraddizione

$$_{a,b}I_{p'} > 1\ ;\ _{b,c}I_{p'} > 1\ ;\ _{a,c}I_{p'} < 1$$

2° *Caso*: $_{a,b}I_p > 1$; $_{b,c}I_p > 1$; $_{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p\,(1+\varepsilon) > 1$; con $0 < \varepsilon$

prendendo

$$p'_{ai} = \frac{p_{ai}}{(1+\varepsilon_1)\ _{a,b}I_p}\ ;\ p'_{ci} = p_{ci}\,(1+\varepsilon_2)\ _{b,c}I_p\ ;\ (i = 1, 2, \ldots s)$$

con $\varepsilon_1, \varepsilon_2 > 0$ e procedendo in modo analogo al 1° caso avremo :

$$ {}_{a,b}I_{p'} = \frac{{}_{a,b}I_p}{(1+\varepsilon_1)\,{}_{a,b}I_p} = \frac{1}{1+\varepsilon_1} < 1 $$

$$ {}_{b,c}I_{p'} = \frac{{}_{b,c}I_p}{(1+\varepsilon_2)\,{}_{b,c}I_p} = \frac{1}{1+\varepsilon_2} < 1 $$

$$ {}_{a,c}I_{p'} = \frac{{}_{a,c}I_p}{(1+\varepsilon_1)(1+\varepsilon_2)\,{}_{a,b}I_p \cdot {}_{b,c}I_p} = \frac{(1+\varepsilon)\,{}_{a,b}I_p \cdot {}_{b,c}I_p}{(1+\varepsilon_1)(1+\varepsilon_2)\,{}_{a,b}I_p \cdot {}_{b,c}I_p} = $$

$$ = \frac{1+\varepsilon}{(1+\varepsilon_1)(1+\varepsilon_2)} $$

ora, affinchè sia ${}_{a,c}I_{p'} > 1$ basta prendere

$$ \varepsilon_1 = \varepsilon_2 = \frac{\varepsilon}{n} \quad \text{dove } n > 1 + \sqrt{1+\varepsilon} $$

infatti

$$ {}_{a,c}I_{p'} = \frac{1+\varepsilon}{\left(1+\frac{\varepsilon}{n}\right)^2} = \frac{1+\varepsilon}{1+\frac{\varepsilon}{n}\left(2+\frac{\varepsilon}{n}\right)} > 1 \text{ per } n > 1 + \sqrt{1+\varepsilon} $$

abbiamo così per certi (p'_{ai}) e (p'_{ci}) la contraddizione

$$ {}_{a,b}I_{p'} < 1 \;.\; {}_{b,c}I_{p'} < 1 \quad \text{e} \quad {}_{a,c}I_{p'} > 1 $$

con gli stessi (p_{bi}), (q_{ai}), (q_{bi}), (q_{ci}).

3° *Caso* : ${}_{a,b}I_p < 1$; ${}_{b,c}I_p < 1$; ${}_{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p\,(1-\varepsilon) < 1$

prendendo :

$$ p'_{ai} = \frac{p_{ai}}{(1-\varepsilon_1)\,{}_{a,b}I_p} ; \; p'_{ci} = p_{ci}\,(1-\varepsilon_2)\,{}_{b,c}I_p \; ; \; (i = 1, 2, \ldots . s) $$

con $0 < \varepsilon_1, \varepsilon_2, < 1$ avremo :

$$ {}_{a,b}I_{p'} = \frac{1}{1-\varepsilon_1} > 1 \; ; \; {}_{b,c}I_{p'} = \frac{1}{1-\varepsilon_2} > 1 $$

$$ {}_{a,c}I_{p'} = \frac{1-\varepsilon}{(1-\varepsilon_1)(1-\varepsilon_2)} < 1 \quad \text{per} \quad \varepsilon_1 = \varepsilon_2 = \frac{\varepsilon}{2} $$

Si giunge così alla contraddizione:

$$_{a,b}I_{p'} > 1 \; ; \; _{b,c}I_{p'} > 1 \; ; \; _{a,c}I_{p'} < 1$$

4° *Caso*: $_{a,b}I_p < 1 \; ; \; _{b,c}I_p < 1 \; ; \; _{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p \, (1+\varepsilon) < 1 \quad (\varepsilon > 0)$

prendendo come al solito

$$p'_{ai} = \frac{p_{ai}}{(1+\varepsilon_1)\, {}_{a,b}I_p} \; ; \; p'_{ci} = p_{ci}\,(1+\varepsilon_2)\, {}_{b,c}I_p \quad \text{con } \varepsilon_1, \varepsilon_2 > 0$$

avremo:

$$_{a,b}I_{p'} = \frac{1}{1+\varepsilon_1} < 1 \; ; \; _{b,c}I_{p'} = \frac{1}{1+\varepsilon_2} < 1$$

$$_{a,c}I_{p'} = \frac{1+\varepsilon}{(1+\varepsilon_1)\,(1+\varepsilon_2)} > 1 \text{ se } \varepsilon_1 = \varepsilon_2 = \frac{\varepsilon}{n} \quad \text{con} \quad n > 1 + \sqrt{1+\varepsilon}$$

giungendo alla contraddizione:

$$_{a,b}I_{p'} < 1 \; ; \; _{b,c}I_{p'} < 1 \; ; \; _{a,c}I_{p'} > 1$$

5° *Caso*: $_{a,b}I_p < 1 \; ; \; _{b,c}I_p > 1 \; ; \quad _{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p \, (1-\varepsilon) \lesseqgtr 1$

se $_{a,c}I_p < 1$ possiamo prendere

$$p'_{ai} = \frac{p_{ai}}{(1-\varepsilon_1)\, {}_{a,b}I_p} \; ; \; p'_{ci} = p_{ci}\,(1-\varepsilon_2)\, {}_{b,c}I_p$$

per cui

$$_{a,b}I_{p'} = \frac{1}{1-\varepsilon_1} > 1 \; ; \; _{b,c}I_{p'} = \frac{1}{1-\varepsilon_2} > 1$$

e basterà prendere $\varepsilon_1 = \varepsilon_2 = \dfrac{\varepsilon}{2}$ per rendere

$$_{a,c}I_{p'} = \frac{1-\varepsilon}{\left(1-\dfrac{\varepsilon}{2}\right)^2} = \frac{1-\varepsilon}{1-\varepsilon+\dfrac{\varepsilon^2}{4}} < 1$$

giungendo alla contraddizione:

$$_{a,b}I_{p'} > 1 \; ; \; _{b,c}I_{p'} > 1 \text{ e } _{a,c}I_{p'} < 1$$

Se ${}_{a,c}I_p > 1$ otterremo la stessa contraddizione seguendo la via identica, cioè ponendo :

$$p'_{ai} = \frac{p_{ai}}{(1 - \varepsilon_1)\ {}_{a,b}I_p} ; p'_{ci} = p_{ci}\,(1 - \varepsilon_2)\ {}_{b,c}I_p \text{ ecc.}$$

6° *Caso*: ${}_{a,b}I_p < 1$; ${}_{b,c}I_p > 1$; ${}_{a,c}I_p = {}_{a,b}I_p \cdot {}_{b,c}I_p\ (1 + \varepsilon) \lesseqgtr 1$

Ambedue i sottocasi ${}_{a,c}I_p \lesseqgtr 1$ saranno ricondotti alla contraddizione ${}_{a,b}I_{p'} < 1$; ${}_{b,c}I_{p'} < 1$; ${}_{a,c}I_{p'} > 1$ ponendo :

$$p'_{ai} = \frac{p_{ai}}{(1 + \varepsilon_1)\ {}_{a,b}I_p} ; p'_{ci} = p_{ci}\,(1 + \varepsilon_2)\ {}_{b,c}I_p$$

e prendendo poi come al solito $\varepsilon_1 = \varepsilon_2 = \frac{\varepsilon}{n}$ con $n > 1 + \sqrt{1 + \varepsilon}$

Rimane così completamente dimostrato quanto asserito nell'enunciato del teorema.

Da notare che il teorema vale ovviamente per le formule non circolari del Paasche e Laspeyres ; per convincersi di ciò basta ripetere le stesse dimostrazioni per le espressioni generali :

$${}_{x,y}I_p = F\,(p_x\,,\,p_y\,;\,q_x) \text{ oppure } {}_{x,y}I_p = F\,(p_x\,,\,p_y\,;\,q_y)$$

Le stesse considerazioni e conclusioni del teorema valgono anche rispetto agli indici del potere di acquisto ottenuti mediante le suddette formule non circolari ; infatti anch'essi si possono considerare da un punto di vista matematico come indici dei prezzi se gli indici dei prezzi da cui derivano soddisfano alla proprietà di reversibilità rispetto ai tempi e luoghi.

Infatti posto $\frac{S_a}{S_b}$ l'indice dei salari del tempo o luogo a rispetto al tempo o luogo b, allora l'indice del potere di acquisto come già è stato ricordato sarà :

$${}_{a,b}I_A = \frac{\frac{S_a}{S_b}}{{}_{a,b}I_p}$$

da cui per le propr. di reversibilità citata

$$\left({}_{a,b}I_p = \frac{1}{{}_{b,a}I_p}\right)$$

abbiamo :

$${}_{a,b}I_A = {}_{b,a}I_p \cdot \frac{S_a}{S_b} = {}_{b,a}I_{p''}$$

dove $p''_{ai} = p_{ai} \cdot S_a$ oppure $p''_{ai} = \frac{p_{ai}}{S_b}$

dove $p''_{bi} = p_{bi} \cdot S_b$ oppure $p''_{bi} = \frac{p_{bi}}{S_a}$

Cioè un indice dei prezzi p'' di b rispetto ad a, con ciò si ricade immediatamente nel teorema precedente.

3. — *Condizioni sufficienti affinchè un indice sia circolare.* Data una successione di n tempi o luoghi : $a_1, a_2 \ldots a_n$, affinchè si abbia un indice generale dei prezzi che soddisfi per questi n tempi o luoghi alla proprietà circolare, basta imporre ad esso di essere della forma :

$$I_{a_i, a_j} = \frac{F\,(P_{a_i,t}; \lambda_1, \lambda_2 \ldots \lambda_m)}{F\,(P_{a_j,t}; \lambda_1, \lambda_2 \ldots \lambda_m)} \quad \text{con } t = 1, 2, \ldots k \qquad (1)$$

con F = funzione arbitraria
con $P_{a_i,t}$ = prezzo del bene t-esimo nel tempo o luogo a_i
con $P_{a_j,t}$ = » » » » » a_j
e $\lambda_1, \lambda_2 \ldots \lambda_m$ sono i parametri fissi (il perchè lo vedremo subito) relativi all'*n-pla* : $a_1, a_2, \ldots, a_n$

Infatti :

$$I_{a_i, a_j} = \frac{F\,(P_{a_i,t}; \lambda_1, \ldots, \lambda_m)}{F\,(P_{a_k,t}; \lambda_1, \ldots, \lambda_m)} \cdot \frac{F\,(P_{a_k,t}; \lambda_1, \ldots \lambda_m)}{F\,(P_{a_l,t}; \lambda_1, \ldots \lambda_m)} \cdots$$

$$\cdots \frac{F\,(P_{a_s,.}; \lambda_1, \ldots \lambda_m)}{F\,(P_{a_z,t}; \lambda_1, \ldots \lambda_m)} \cdot \frac{F\,(P_{a_z,t}; \lambda_1, \ldots \lambda_m)}{F\,(P_{a_j,t}; \lambda_1, \ldots \lambda_m)}$$

dove cioè il permanere nella (1) della stessa funzione F ed egli stessi parametri per ogni coppia (a_i, a_j) conduce alla proprietà circolare.

È naturale che i parametri $\lambda_1, \lambda_2, \ldots \lambda_m$ sono in generale prezzi e quantità; e siccome per soddisfare alla proprietà circolare con la (1) essi devono costituire un sistema rigido invariante rispetto a tutte le coppie di confronto possibili, a meno che non ci siano particolari ragioni economiche per scegliere i prezzi e quantità di uno solo o di alcuni degli n tempi o luoghi dovremo includervi e prezzi e quantità di tutti gli n tempi e luoghi.

Nella « formula ideale » ad esempio non si verifica la proprietà circolare perchè permane la stessa funzione F ma non gli stessi parametri e cioè in essa:

$$I_{a_i,a_j} = \frac{F(P_{a_i,t}; Q_{a_i,t}, Q_{a_j,t})}{F(P_{a_j,t}; Q_{a_i,t}, Q_{a_j,t})}; \quad I_{a_k,a_i} = \frac{F(P_{a_k,t}; Q_{a_k,t}, Q_{a_i,t})}{F(P_{a_i,t}; Q_{a_k,t}, Q_{a_i,t})}$$

per la coppia (a_i, a_j) ci sono i parametri $Q_{a_i,t}, Q_{a_j,t}$ mentre per la coppia (a_k, a_i) si sono i parametri $Q_{a_k,t}, Q_{a_i,t}$.

Riprendendo la (1) possiamo dire in generale che per l'n-pla $a_1, a_2, \ldots a_n$ l'espressione generale generatrice di indici circolari può scriversi:

$$I_{a_i,a_j} = \frac{F(P_{a_i,t}; Q_{s,t}; P_{s,t})}{F(P_{a_j,t}; Q_{s,t}; P_{s,t})} \quad \text{dove} \quad \begin{matrix} s = a_1, a_2, \ldots a_n \\ t = 1, 2, \ldots k \end{matrix}$$

e più precisamente:

$$I_{a_i,a_j} = \frac{\Omega[\varphi(P_{a_i,t}; Q_{s,t}; P_{s,t})]}{\Omega[\varphi(P_{a_j,t}; Q_{s,t}; P_{s,t})]} \qquad (2)$$

dove Ω e φ sono due funzioni od operatori che verranno definiti di volta in volta.

Se $\varphi_{s,i} = \sum_t P_{a_i,t} \cdot Q_{s,t}$ e $\Omega =$ Media aritmetica semplice rispetto all'indice s abbiamo:

$$I_{a_i,a_j} = \frac{\frac{1}{n}\sum_s \sum_t P_{a_i,t} Q_{s,t}}{\frac{1}{n}\sum_s \sum_t P_{a_j,t} Q_{s,t}} = \frac{\sum P_{a_i,t} \frac{\sum_s Q_{s,t}}{n}}{\sum P_{a_j,t} \frac{\sum_s Q_{s,t}}{n}}$$

che è la (I) del Gini

mantenendo la stessa φ_s e facendo Ω = media geometrica rispetto all'indice s avremo:

$$I_{a_i,a_j} = \frac{\sqrt[n]{\pi_s \sum_t P_{a_i,t} Q_{s,t}}}{\sqrt[n]{\pi_s \sum_t P_{a_j,t} Q_{s,t}}} = \sqrt[n]{\frac{\sum_t P_{a_i,t} Q_{s,t}}{\pi_s \sum_t P_{a_j,t} Q_{s,t}}}$$

che è la (III) del Gini

Se facciamo $\varphi_{s,i} = \sum_t P_{a_i,t} \frac{Q_{s,t} P_{s,t}}{\sum_s P_{s,t}}$ e Ω è la media aritmetica semplice rispetto all'indice s avremo:

$$I_{a_i,a_j} = \frac{\frac{1}{n} \sum_s \sum_t P_{a_i,t} \frac{Q_{s,t} P_{s,t}}{\sum_s P_{s,t}}}{\frac{1}{n} \sum_s \sum_t P_{a_j,t} \frac{Q_{s,t} P_{s,t}}{\sum_s P_{s,t}}} = \frac{\sum_t P_{a_i,t} \frac{\sum_s P_{s,t} Q_{s,t}}{\sum_s P_{s,t}}}{\sum P_{a_j,t} \frac{\sum_s P_{s,t} Q_{s,t}}{\sum_s P_{s,t}}}$$

che è la (II) del Gini.

Se $\varphi_{s,i}$ è *un indice* non circolare di a_i rispetto ad s ($s = a_1, a_2, \ldots a_n$) e Ω è la media geometrica di tali n indici che si ottengono confrontando a numeratore della (2) a_i con gli altri n tempi o luoghi compreso a_i stesso ed a numeratore della (2) a_j con gli altri n tempi o luoghi abbiamo ad esempio se φ è la « formula ideale »:

$$\varphi_{s,i} = \sqrt{\frac{\sum_t P_{a_i,t} Q_{s,t}}{\sum_t P_{s,t} Q_{s,t}} \frac{\sum_t P_{a_i,t} Q_{a_i,t}}{\sum_t P_{s,t} Q_{a_i,t}}} = \underline{a_i s}$$

prendendo la media geometrica (applicando cioè Ω) rispetto all'indice s avremo:

$$I_{a_i,a_j} = \frac{\sqrt[u]{\underline{a_i a_1} \cdot \underline{a_i a_2} \ldots \underline{a_i a_i} \ldots \underline{a_i a_n}}}{\sqrt[n]{\underline{a_j a_1} \cdot \underline{a_j a_2} \ldots \underline{a_j a_j} \ldots \underline{a_j a_n}}} = \sqrt[n]{\frac{\underline{a_i a_1}}{\underline{a_j a_1}} \frac{\underline{a_i a_2}}{\underline{a_j a_2}} \frac{\underline{a_i a_3}}{\underline{a_j a_3}} \cdots \frac{\underline{a_i a_n}}{\underline{a_j a_n}}}$$

che è la (IV) del Gini.

Altre se ne possono trarre: ad esempio se facciamo ancora $\varphi_{s,t} = \Sigma_t P_{a_i,t} Q_{s,t}$ e $\Omega =$ media armonica semplice rispetto all'indice s avremo:

$$I_{a_i,a_j} = \frac{\Sigma_s \frac{1}{\Sigma_t P_{a_i,t} Q_{s,t}}}{\Sigma_s \frac{1}{\Sigma_t P_{a_j,t} Q_{s,t}}} = \frac{\Sigma_s \frac{1}{\Sigma_t P_{a_j,t} Q_{s,t}}}{\Sigma_s \frac{1}{\Sigma_t P_{a_i,t} Q_{s,t}}}$$

Oppure se facciamo

$$\varphi_{1,i} = \Sigma P_{a_i,t} Q_{a_1,t}\,;\ \varphi_{2,i} = \frac{\Sigma_t P_{a_i,t} \sqrt{Q_{a_i} Q_{a_2,t}}}{\Sigma_t P_{a_i,t} Q_{a_1,t}}$$

$$\varphi_{3,i} = \frac{\Sigma_t P_{a_i,t} \sqrt[3]{Q_{a_1 t} \cdot Q_{a_2 t} \cdot Q_{a_3 t}}}{\Sigma_t P_{a_i,t} \sqrt{Q_{a_1,t} \cdot Q_{2,t}}}\,;$$

cioè in generale:

$$\varphi_{s,i} = \frac{\Sigma_t P_{a_i,t} \sqrt[s]{\pi_{l=1}^{s} Q_{a_l,t}}}{\Sigma_t P_{a_i,l} \sqrt[s-1]{\pi_{l=1}^{s-1} Q_{a_l,t}}}$$

e facendo $\Omega = \pi_s^n$ (simbolo di prodotto) avremo:

$$I_{a_i,a_j} = \frac{\Sigma_t P_{a_i,t} \sqrt[n]{Q_{a_1,t} \cdot Q_{a_2,t} \cdots Q_{a_n,t}}}{\Sigma_t P_{a_j,t} \sqrt[n]{Q_{a_1,t} \cdot Q_{a_2,t} \cdots Q_{a_n,t}}}$$

e così via.

A questo punto dobbiamo riconoscere che le 4 formule del Gini costituiscono le particolarizzazioni più notevoli dello schema generale esposto, sia per la semplicità formale sia per il significato economico concreto dei numeratori e denominatori di esse.

Per non arrivare a delle espressioni circolari che presentino il carattere di una semplice curiosità algebrica occorre evidentemente imporre opportune condizioni a φ come ad esempio che sia sempre :

$$\Delta_P \varphi (P_{a_i}, \lambda_1 \lambda_2, \ldots \lambda_m) > O$$

cioè che φ sia crescente con il rispettivo P_{a_i}, e così via ...

4. — *Considerazioni sullo schema di Ragnar Frisch.* Il Frisch in un suo lavoro già citato (1) enumera per gli indici dei prezzi le 7 proprietà seguenti : (diverse delle quali già ricordate più indietro) :

1. *Proprietà di identità* (Identity test) - cioè $I_{a,a} = 1$

2. *Reversibilità dei tempi o luoghi* (time *oppure* place reversal test) cioè $I_{a,b} = \frac{1}{I_{b,a}}$

3. *Proprietà di indipendenza* - (Base test) cioè : $\frac{I_{a,x}}{I_{b,x}} = R_{a,b}$ che significa imporre al rapporto $\frac{I_{a,x}}{I_{b,x}}$ di essere una funzione di a e b soltanto e quindi indipendente da x.

4. *Proprietà circolare* (circulare test) cioè $I_{a,x} \cdot I_{x,b} \cdot I_{b,a} = 1$; abbiamo visto che la propr. circolare implica sempre le propr. 1 , 2 , 3 ma non è vero il contrario.

5. *Proprietà della Commensurabilità* (Commensurability test). Questa richiede che — se si cambia l'unità di misura di un certo bene ossia se il relativo prezzo p e quantità q vengono sostituiti al tempo o luogo a e b rispettiv. con λp e $\frac{q}{\lambda}$ — lasci invariato $I_{a,b}$

6. *Proprietà di determinatezza* (Determinateness test).

Essa richiede che un indice non divenga nullo, infinito o indeterminato se una singola quantità o singolo prezzo diviene zero

---

(1) R. Frisch - *Necessary and sufficient conditions* ecc. in « Journal of the Amer. Stat. Ass. » Dic. 1930.

7. *Reversibilità dei fattori* (Factor Reversal test)

cioè $\frac{\Sigma P_a Q_a}{\Sigma P_b Q_b} = {}_{a,b}I_p \cdot {}_{a,b}I_q$ dove ${}_{a,b}I_q$ è l'indice delle quantità ottenuto *scambiando in* ${}_{a,b}I_p$ *i prezzi con le quantità.*

Ora il Frisch partendo dall'espressione generale:

$$I_{a_i, a_j} = f(P_{a_i,t}, P_{a_j,t}; Q_{a_i,t}, Q_{a_j,t}) \qquad (3)$$

che rappresenta un generico indice dei prezzi del tempo a_i rispetto al tempo a_j (lui parla sempre e solo di tempi) e dove come al solito i P e Q sono prezzi e quantità, imponendo successivamente alla (3) le proprietà elencate giunge a degli indici che soddisfano solo a coppie delle proprietà indipendenti (1) del prospetto precedente e solo in un caso a tre proprietà indipendenti simultaneamente: si tratta in questo caso della formula:

$$I_{a_i, a_j} = \pi_t \left(\frac{P_{a_i,t}}{P_{a_j,t}}\right)^{c_t} \sqrt{\frac{\Sigma_t P_{a_i,t} Q_{a_i,t}}{\Sigma_t P_{a_j,t} Q_{a_j,t}} \Big/ \pi_t \frac{P_{a_i,t} Q_{a_i,t}}{P_{a_j,t} Q_{a_j,t}}}$$

che concilia le proprietà 4, 5, 7.

Inoltre *esso conclude che non esistono indici* (*che posseggono derivate parziali rispetto ai P e Q*) *che soddisfano simultaneamente alle proprietà:*

(γ) { *Circolare* (circular test)
di Commensurabilità (Commensurability test)
di Determinatezza (Determinateness test) (2)

Ora vediamo subito che le 4 formule del Gini ed altre che discendono dal nostro schema esposto al paragr. 3 della parte metodol. *dimostrano invece il contrario. Più precisamente che le*

---

(1) Poichè la proprietà circolare implica sempre la 1, 2 e 3

(2) Vedi anche: R. Frisch - *Annual Survey of General Economic Theory: The Problem of Index Numbers* in « Econometrica » 1936 N. 1, pagina 7.

formule del Gini soddisfano rispettivamente: la (I), (II) e (III) alle prime 6 proprietà enumerate e quindi simultaneamente alle (γ) e la (IV) soddisfa simultaneamente addirittura a tutte le 7 proprietà enumerate dal Frisch possedendo inoltre derivate parziali rispetto a P e Q.

Per dimostrare ciò basterà prendere l'espressione generale del nostro schema che comprende le 4 form. del Gini e cioè

$$I_{a_i,a_j} = \frac{\Omega\left[\varphi\left(P_{a_i,t}; Q_{s,t}; P_{s,t}\right)\right]}{\Omega\left[\varphi\left(P_{a_j,t}; Q_{s,t}; P_{s,t}\right)\right]}$$

Anzitutto si soddisfa alla proprietà circolare (per le note ragioni del paragrafo 3 già citato) e quindi anche alle 3 proprietà implicite in questa e cioè la 1, 2, 3 della enumerazione.

Prendendo successivamente:

$$\varphi_{s,i} = \sum_t P_{a_i,t}\, Q_{s,t}$$

$$\varphi_{s,i} = \sum_t P_{a_i,t} \frac{P_{s,t}\, Q_{s,t}}{\sum_s P_{s,t}}$$

$$\varphi_{s,i} = \sqrt{\frac{\sum_t P_{a_i,t}\, Q_{s,t}}{\sum_t P_{s,t}\, Q_{s,t}} \frac{\sum_t P_{a_i,t}\, Q_{a_i,t}}{\sum P_{s,t}\, Q_{a_i,t}}}$$

che conducono alle 4 formule del Gini (scegliendo nel modo che sappiamo Ω) si soddisfa alla proprietà di *Commensurabilità*; infatti (ricordando le notazioni del paragrafo 3) se in ognuno dei k beni considerati viene alterata l'unità di misura, ingrandendola ad esemp. di λ_t volte nel bene generico t allora le quantità di questo bene in tutti i diversi tempi e luoghi che figurano in ogni φ_{si} $(s = a_1, a_2 \ldots a_n)$ verranno moltiplicate per $\frac{1}{\lambda_t}$ ed i relativi prezzi per λ_t. Questa operazione lascia invariate come si vede subito le espressioni:

$$\varphi_{s,i}, \lambda_t = \sum_t \lambda_t P_{a_i,t} \frac{Q_{s,t}}{\lambda_t} = \sum_t P_{a_i,t}\, Q_{s,t} = \varphi_{s,i}$$

$$\varphi_{s,i}, \lambda_t = \sum \lambda_t P_{a_i,t} \frac{\lambda_t P_{s,t} Q_{s,t}/\lambda_t}{\lambda_t \sum_s P_{s,t}} = \sum_t P_{a_i,t} \frac{P_{s,t}\, Q_{s,t}}{\sum_s P_{s,t}} = \varphi_{s,i}$$

$$\varphi_{s,t}\,,\ \lambda_t = \sqrt{\frac{\sum_t \lambda_t P_{a_i,t}\, Q_{s,t}/\lambda_t \ \sum_t \lambda_t P_{a_i,t}\, Q_{a_i,t}/\lambda_t}{\sum_t \lambda_t P_{s,t}\, Q_{s,t}/\lambda_t \ \sum_t \lambda_t P_{s,t}\, Q_{a_i},/\lambda_t}} = \varphi_{s,t}$$

Si soddisfa alla proprietà di *determinatezza* perchè le varie $\varphi_{s,t}$ si mantengono finite, determinate e non si annullano annullando un singolo prezzo o quantità poichè in tal caso si annulla solo un termine in alcune somme di più termini.

Inoltre le tre espressioni di $\varphi_{s,t}$ che conducono alle 4 formule del Gini, posseggono derivate parziali rispetto ai P e Q e quindi anche $\dfrac{\Omega\,(\varphi_i)}{\Omega\,(\varphi_j)} = I_{i,j}$ data la natura di Ω che abbiamo scelto e che possiamo sempre scegliere in modo opportuno.

Infine la (IV) del Gini soddisfa pure alla *proprietà di reversibilità rispetto ai fattori* in quanto: (vedi più indietro la definizione della pag. 7).

$$a_i a_j I_p \cdot a_i a_j I_q = \frac{\sum P_{a_i} Q_{a_i}}{\sum P_{a_j} Q_{a_j}}$$

come è facilissimo verificare.

Il perchè delle affermazioni del Frisch, e perchè non valgono per le formule del Gini è semplice.

Il Frisch è arrivato a tali conclusioni partendo dall'espressione generale (3) che non è la sola possibile poichè infatti essa non comprende la impostazione cui derivano le formule del Gini che è poi la nostra impostazione del paragrafo 3, perciò vedremo subito che il Frisch non poteva arrivare altro che ai risultati cui è giunto.

Egli doveva quindi riferire esplicitamente le sue conclusioni alla sua particolare impostazione del problema che non è la sola possibile.

Esaminiamo ora le due diverse formule generali di partenza:

$$I_{a_i a_j} = f\,(P_{a_i,t}\,,\ P_{a_j,t}\,,\ Q_{a_i,t}\,,\ Q_{a_j,t})\ ;\ I_{a_i a_j} = f\,(P_{a_i,t}\,P_{a_j,t}\,;\lambda_1\,,\lambda_2\,,\ldots,\lambda_m)$$

Nella prima (Frisch) figurano solo i prezzi e quantità relativi alla coppia $(a_i\,,\ a_j)$ mentre nell'altra (la nostra che com-

prende le formule del Gini) figurano i prezzi di a_i ed a_j e un numero di parametri o coppie di parametri uguale al numero dei tempi o luoghi per cui interessa che l'indice abbia certe proprietà.

Fissato così il problema concreto consistente nello stabilire il numero n dei tempi e luoghi che interessa far percorrere allo indice, fissate certe proprietà di cui vogliamo e possiamo munire l'indice quando esso passa per qualunque coppia di tempi e luoghi fra gli n scelti, si può costruire l'indice adatto a ciò con lo schema cui appartengono le formule del Gini, che conduce a soddisfare in generale *sempre e solo* per gruppi di coppie scelte fra le $\binom{n+1}{2}$ coppie possibili (combinazioni con ripetizioni) le proprietà cui è stato possibile conferire all'indice, poichè come abbiamo già detto n sono i tempi e luoghi che interessano.

Il Frisch invece si disinteressa delle dimensioni del problema concreto e giunge ad indici che soddisfano proprietà in modo indefinito (in quanto la sua espressione generale non ha parametri ancorati al particolare problema) e quindi tali proprietà vengono soddisfatte anche per tempi e luoghi che non interessano affatto gli n del problema concreto (n = naturale qualunque).

Quindi vediamo che imponendo alle due formule generali una stessa condizione (che sia possibile soddisfare ad ambedue) si giunge ad un diverso modo di soddisfarvi.

Le formule del Gini vi soddisfano perfettamente ma nel modo più economico dal punto di vista del numero dei tempi e luoghi per cui interessa che la proprietà valga, mentre quelle del Frisch possiedono tale proprietà indefinitamente anche cioè riguardo a tempi o luoghi estranei al problema concreto.

In conclusione il fatto del minor numero di certe proprietà indipendenti soddisfatto simultaneamente da indici del Frisch in confronto a quelli del Gini, e l'impossibilità di conciliare certe terne di proprietà in un indice come succede al Frisch a differenza del Gini, sta a dimostrare che la maggiore economicità (nel senso già spiegato) che consentono le formule del Gini ed altre analoghe del nostro schema nel soddisfare una certa proprietà, si ripercuote favorevolmente riguardo al soddisfacimento di un maggior numero di proprietà indispensabili.

Note pag. 185 186

$$ {}_{a,b}I_t = \sqrt[7]{\prod_{k=1}^{7} \frac{t_{a,k}}{t_{b,k}}} = $$

$$ = \frac{S_b}{S_a} \sqrt[7]{\frac{\sum P_{i,a} Q_{i,a}}{\sum P_{i,b} Q_{i,a}} \frac{\sum P_{i,a} Q_{i,b}}{\sum P_{i,b} Q_{i,b}} \frac{\sum P_{i,a} Q_{i,c}}{\sum P_{i,b} Q_{i,c}} \cdots \frac{\sum P_{i,a} Q_{i,z}}{\sum P_{i,b} Q_{i,z}}} = \frac{S_b}{S_a} {}_{a,b}I_p ; $$

$$ {}_{a,b}I_A = \frac{S_b}{S_a} \frac{1}{{}_{a,b}I_p} = \frac{1}{{}_{a,b}I_t} $$

$t_{a,k} = \dfrac{\sum P_{i,a} Q_{i,k}}{S_a}$ = tempo impiegato a percepire con il salario orario S_a l'importo $\sum P_{i,a} Q_{i,k}$ reilativo alla razione della città K con i prezzi della città a

S_a = Salario orario della città a ; S_b = Salario della città b

${}_{a,b}I_t$ = indice dei tempi della citta a rispetto alla città b .

${}_{a,b}I_p$ = indice dei prezzi della città a rispetto la (formula III del Gini) (*Successive weights sistem*)

${}_{a,b}I_A$ = indice del potere di acquisto del salario S_a dellq città a rispetto a quello del salario S_b della città b riguardo alle razioni alimentari.

Note pag. 187 188

$$ {}_{a,b}I_t = \frac{S_b}{S_a} \sqrt{\frac{aa}{ba} \frac{ab}{bb} \frac{ac}{bc} \frac{ad}{bd} \frac{ae}{be} \frac{af}{bf} \frac{ag}{bg}} = \frac{S_b}{S_a} {}_{a,b}I_p $$

$$ {}_{a,b}I_A = \frac{1}{{}_{a,b}I_t} = \frac{S_a}{S_b} \frac{1}{{}_{a,b}I_p} $$

dove ${}_{a,b}I_p$ è l'indice dei prezzi della città a rispetto alla città b secondo la formula IV bis del Gini - *circular weight systems*) in cui $\underline{aa}$, $\underline{ba}$, $\underline{ab}$, $\underline{bb}$, ... $\underline{ag}$, $\underline{bg}$ sono indici elementari dei prezzi secondo la formula ideale.

S_a , S_b = rispett. salari della città a e b,

${}_{a,b}I_t$ = indice dei tempi della città a rispetto alla città b ;

${}_{a,b}I_A$ = indice del potere di acquisto del Salario S_a rispetto a S_b relativamente alle razioni alimentari.

TABELLA I[a]

Tabella degli indici I_t, I_P, I_A di ogni città, rispetto a ciascun altra.

			Bolzano				Lecce				Salerno				Sassari			
			I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
Bolzano . .	M	I_t	1, —	1, —	1, —	1, —	1,110	1,302	1,381	1,224	1,085	1,267	1,269	1,172	1,099	1,157	1,460	
		I_P	1, —	1, —	1, —	1, —	1,021	1,021	1,021	1,021	1,041	1,041	1,041	1,041	0,993	0,993	0,993	0,993
		I_A	1, —	1, —	1, —	1, —	0,901	0,768	0,731	0,817	0,922	0,789	0,788	0,853	0,910	0,864	0,685	
	C	I_t	1, —	1, —	1, —	1, —	1,075	1,101	1,124	1,314	1,111	1,164	1,109	91,147	1,015	1,111	0,952	1,148
		I_P	1, —	1, —	1, —	1, —	1,021	1,021	1,021	1,021	1,041	1,041	1,041	1,041	0,993	0,993	0,993	0,993
		I_A	1, —	1, —	1, —	1, —	0,930	0,908	0,890	0,761	0,900	0,859	0,910	0,872	0,985	0,900	1,050	0,871
Lecce . . .	M	I_t	0,901	0,768	0,731	0,817	1, —	1, —	1, —	1, —	0,978	0,974	0,928	0,959	0,990	0,889	1,068	
		I_P	0,980	0,980	0,980	0,980	1, —	1, —	1, —	1, —	1,020	1,020	1,020	1,020	0,983	0,973	0,973	0,973
		I_A	1,110	1,302	1,381	1,224	1, —	1, —	1, —	1, —	1,023	1,027	1,078	1,043	1,010	1,125	0,936	—
	C	I_t	0,930	0,908	0,890	0,761	1, —	1, —	1, —	1, —	1,034	1,058	0,978	0,873	0,944	1,009	0,847	0,874
		I_P	0,980	0,980	0,980	0,980	1, —	1, —	1, —	1, —	1,020	1,020	1,020	1,020	0,973	0,973	0,973	0,973
		I_A	1,075	1,101	1,124	1,314	1, —	1, —	1, —	1, —	0,967	0,945	1,023	1,145	1,059	0,991	1,180	1,144
Salerno. . .	M	I_t	0,922	0,789	0,788	0,853	1,023	1,027	1,078	1,043	1, —	1, —	1, —	1, —	1,013	0,913	1,151	
		I_P	0,961	0,961	0,961	0,961	0,980	0,980	0,980	0,980	1, —	1, —	1, —	1, —	0,954	0,954	0,954	0,954
		I_A	1,085	1,267	1,269	1,172	0,978	0,974	0,928	0,959	1 ,—	1, —	1, —	1, —	0,987	1,095	0,869	
	C	I_t	0,900	0,859	0,910	0,872	0,967	0,945	1,023	1,145	1, —	1, —	1, —	1, —	0,914	0,954	0,867	1,001
		I_P	0,961	0,961	0,961	0,961	0,980	0,980	,0980	0,980	1, —	1, —	1, —	1, —	0,954	0,954	0,954	0,954
		I_A	1,111	1,164	1,099	1,147	1,034	1,058	0,978	0,873	1, —	1, —	1, —	1, —	1,094	1,048	1,154	0,999
Sassari. . .	M	I_t	0,910	0,864	0,685		1,010	1,125	0,936		0,987	1,095	0,869	—	1, —	1, —	1, —	1, —
		I_P	1,007	1,007	1,007	1.007	1,027	1,027	1,027	1,027	1,048	1,048	1,048	1,048	1, —	1, —	1, —	1, —
		I_A	1,099	1,157	1,460		0,990	0,889	1,068		1,013	0,913	1,151		1, —	1, —	1, —	1, —
	C	I_t	0,985	0,900	1,050	0,871	1,059	0,991	1,180	1,144	1,094	1,048	1,154	0,999	1, —	1, —	1, —	1, —
		I_P	1,007	1,007	1,007	1,007	1,027	1,027	1,027	1,027	1,048	1,048	1,048	1,048	1, —	1, —	1, —	1, —
		I_A	1,015	1,111	0,952	1,148	0,944	1,009	0,847	0,874	0,914	0,954	0,867	1,001	1, —	1, —	1, —	1, —
Messina . .	M	I_t	0,877	0,827	0,820	0,899	0,975	1,078	1,122	1,101	0,952	1,050	1,041	1,056	0,965	0,958	1,197	—
		I_P	0,894	0,894	0,894	0,894	0,913	0,913	0,913	0,913	0,931	0,931	0,931	0,931	0,888	0,888	0,888	0,888
		I_A	1,140	1,209	1,220	1,112	1,026	0,928	0,891	0,908	1,050	0,952	0,961	0,947	1,036	1,044	0,935	
	C	I_t	0,796	0,772	0,782	0,774	0,857	0,851	0,880	1,017	0,885	0,900	0,860	0,888	0,809	0,858	0,745	0,888
		I_P	0,894	0,894	0,894	0,894	0,913	0,913	0,913	0,913	0,931	0,931	0,931	0,931	0,888	0,888	0,888	0,888
		I_A	1,256	1,295	1,279	1,292	1,167	1,175	1,136	0,983	1,130	1,111	1,163	1,126	1,236	1,166	1,342	1,136
Campobasso	M	I_t	0,997	0,814	0,851	0,887	1,107	1,061	1,165	1,086	1,082	1,033	1,080	1,041	1,097	0,943	1,244	—
		I_P	0,970	0,970	0,970	0,970	0,990	0,990	0,990	0,990	1,010	1,010	1,010	1,010	0,964	0,964	0,964	0,964
		I_A	1,003	1,229	1,175	1,127	0,903	0,943	0,858	0,921	0,924	0,968	0,926	0,961	0,912	1,060	0,804	
	C	I_t	0,798	0,801	0,772	0,769	0,858	0,882	0,868	1,001	0,887	0,933	0,849	0,882	0,811	0,891	0,736	0,883
		I_P	0,970	0,970	0,970	0,970	0,990	0,990	0,990	0,990	1,010	1,010	1,010	1,010	0,964	0,964	0,964	0,964
		I_A	1,253	1,248	1,295	1,300	1,166	1,134	1,152	0,999	1,127	1,072	1,178	1,134	1,233	1,122	1,359	1,133
Milano. . .	M	I_t	1,031	0,986	1,006	1,157	1,145	1,285	1,378	1,418	1,119	1,250	1,277	1,359	1,134	1,141	1,470	—
		I_P	0,940	0,940	0,940	0,940	0,960	0,960	0,960	0,960	0,979	0,979	0,979	0,979	0,934	0,934	0,934	0,934
		I_A	0,970	1,014	0,994	0,864	0,873	0,778	0,726	0,705	0,894	0,800	0,783	0,736	0,882	0,876	0,680	
	C	I_t	1,011	1,007	1,104	1,157	1,088	1,110	1,242	1,521	1,124	1,173	1,214	1,327	1,027	1,119	1,052	1,328
		I_P	0,940	0,940	0,940	0,940	0,960	0,960	0,960	0,960	0,979	0,979	0,979	0,979	0,934	0,934	0,934	0,934
		I_A	0,989	0,993	0,906	0,864	0,919	0,901	0,982	0,657	0,890	0,853	0,824	0,754	0,974	0,894	0,951	0,753

Le città scritte nella testata, orizzontale stanno a numeratore.

Messina,				Campobasso				Milano			
I	II	III	IV	I	II	III	VI	I	II	III	IV
1,140	1,209	1,220	1,112	1,003	1,229	1,175	1,127	0,970	1,014	0,994	0,864
1,118	1,118	1,118	1,118	1,031	1,031	1,031	1,031	1,063	1,063	1,063	1,063
0,877	0,827	0,820	0,899	0,997	0,814	0,851	0,887	1,031	0,986	1,006	1,157
1,256	1,295	1,279	1,292	1,253	1,248	1,295	1,300	0,989	0,993	0,906	0,864
1,118	1,118	1,118	1,118	1,031	1,031	1,031	1,031	1,063	1,063	1,063	1,063
0,796	0,772	0,782	0,774	0,798	0,801	0,772	0,769	1,011	1,007	1,104	1,157
1,026	0,928	0,891	0,908	0,903	0,943	0,858	0,921	0,873	0,778	0,726	0,705
1,096	1,096	1,096	1,096	1,010	1,010	1,010	1,010	1,042	1,042	1,042	1,042
0,985	1,078	1,122	1,101	1,107	1,061	1,165	1,086	1,145	1,285	1,378	1,418
1,167	1,175	1,136	0,983	1,166	1,134	1,152	0,999	0,919	0,901	0,982	0,657
1,096	1,096	1,096	1,096	1,010	1,010	1,010	1,010	1,042	1,042	1,042	1,042
0,857	0,851	0,880	1,017	0,858	0,882	0,868	1,001	1,088	1,110	1,242	1,521
1,030	0,952	0,961	0,947	0,924	0,968	0,926	0,961	0,894	0,800	0,783	0,736
1,074	1,074	1,074	1,074	0,990	0,990	0,990	0,990	1,021	1,021	1,021	1,021
0,952	1,050	1,041	1,056	1,082	1,033	1,080	1,041	1,119	1,250	1,277	1,359
1,130	1,111	1,163	1,126	1,127	1,072	1,178	1,134	0,896	0,853	0,824	0,754
1,074	1,074	1,074	1,074	0,990	0,990	0,990	0,990	1,021	1,021	1,021	1,021
0,885	0,900	0,860	0,888	0,887	0,933	0,849	0,882	1,124	1,173	1,214	1,327
1,036	1,044	0,835		0,912	1,060	0,804		0,882	0,876	0,680	
1,126	1,126	1,126	1,126	1,038	1,038	1,038	1,038	1,070	1,070	1,070	1,070
0,965	0,958	1,197		1,097	0,943	1,244		1,134	1,141	1,470	
1,236	1,166	1,342	1,136	1,233	1,122	1,359	1,133	0,974	0,894	0,951	0,753
1,126	1,126	1,126	1,126	1,038	1,038	1,038	1,038	1,070	1,070	1,070	1,070
0,809	0,858	0,745	0,888	0,811	0,891	0,736	0,883	1,027	1,119	1,052	1,328
1, —	1, —	1, —	1, —	0,880	1,016	0,963	1,014	0,850	0,839	0,814	0,776
1, —	1, —	1, —	1, —	0,922	0,922	0,922	0,922	0,951	0,951	0,951	0,951
1, —	1, —	1, —	1, —	1,137	0,984	1,038	0,986	1,176	1,192	1,228	1,288
1, —	1, —	1, —	1, —	0,998	0,964	1,149	1,007	0,787	0,767	0,708	0,668
1, —	1, —	1, —	1, —	0,922	0,922	0,922	0,922	0,951	0,951	0,951	0,951
1, —	1, —	1, —	1, —	1,002	1,037	0,987	0,993	1,270	1,304	1,412	1,496
1,137	0,984	1,038	0,986	1, —	1, —	1, —	1, —	0,966	0,826	0,845	0,765
1,085	1,085	1,085	1,085	1, —	1, —	1, —	1,	1,031	1,031	1,031	1,031
0,880	1,016	0,963	1,014	1, —	1,	1, —	1, —	1,035	1,211	1,183	1,307
1,002	1,037	0,987	0,993	1, —	1, —	1, —	1, —	0,789	0,750	0,699	0,664
1,085	1,085	1,085	1,085	1, —	1, —	1, —	1,	1,031	1,031	1,031	1,031
0,998	0,964	1,149	1,007	1, —	1, —	1, —	1, —	1,268	1,258	1,431	1,507
1,176	1,192	1,228	1,288	1,035	1,211	1,183	1,307	1, —	1, —	1, —	1, —
1,052	1,052	1,052	1,052	0,970	0,970	0,970	0,970	1, —	1, —	1, —	1, —
0,850	0,839	0,814	0,776	0,966	0,826	0,845	0,765	1, —	1, —	1, —	1, —
1,270	1,304	1,412	1,496	1,268	1,258	1,431	1,507	1, —	1, —	1, —	1, —
1,052	1,052	1,052	1,052	0,970	0,970	0,970	0,970	1, —	1, —	1, —	1, —
0,787	0,767	0,708	0,668	0,789	0,750	0,699	0,664	1, —	1, —	1, —	1, —

M = operai metalmeccanici

C = operai edili

I, II, III, IV indicano rispettivamente :

I = operai specializzati
II = operai qualificati
III = manovali specializzati
IV = manovali comuni

$$ {}_{a,b}I_A = \frac{s_b}{s_a} \frac{\sum_l P'_{al} \frac{\sum_{i=1}^{7} Q_{il}}{7}}{\sum_l P_{bl} \frac{\sum_{i=1}^{7} Q_{il}}{7}} = \frac{s_s}{s_a} \cdot {}_{a,b}I_p $$

$$ {}_{a,b}I_A = \frac{\text{I}}{{}_{a,b}I_t} = \frac{s_a}{s_b} \cdot \frac{\text{I}}{{}_{a,b}I_p} $$

dove :

${}_{a,b}I_p$ = indice dei prezzi della città a rispetto alla città b (formula I del Gini - *average weights system*)

s_a , s_b = salari orari delle città a e b.

${}_{a,b}I_t$ = indice dei tempi della città a rispetto alla città b

${}_{a,b}I_A$ = indice del potere di acquisto del salario s_a rispetto a s_b riguardo alle razioni alimentari.

TABELLA 2ª

Tabella degli indici I_t, I_P, I_A di ogni città rispetto a ciascun altra

		Bolzano				Lecce				Salerno				Sassari			
		I	II	III	IV	I	II	III	AI	I	II	III	IV	I	II	III	IV
Bolzano	M I_t	1, —	1, —	1, —	1, —	1.038	1.218	1.280	1.145	1.036	1.211	1.212	1.121	1.049	1.105	1.395	
	M I_P	1, —	1, —	1, —	1, —	0,955	0,955	0,955	0,955	0,994	0,994	0,994	0,994	0,948	0,948	0,948	0,948
	M I_A	1, —	1. —	1. —	1, —	0,963	0,821	0,781	0,873	0,965	0,826	0,825	0,892	0,953	0,905	0,717	
	C I_t	1, —	1, —	1, —	1, —	1,005	1,031	1,052	1,229	1,060	1,112	1,049	1,095	0,969	1,060	0,909	1.095
	C I_P	1, —	1, —	1, —	1, —	0,955	0,955	0,955	0,955	0,994	0,994	0,994	0,994	0,948	0,948	0,948	0,948
	C I_A	1, —	1, —	1, —	1, —	0,995	0,970	0,951	0,814	0,943	0,899	0,953	0,913	1,032	0,943	1,100	0,913
Lecce	M I_t	0,963	0,821	0,781	0,873	1, —	1, —	1, —	1, —	0,998	0,994	0,947	0,978	1,055	0,907	1,089	
	M I_P	1,048	1,048	1,048	1,048	1, —	1, —	1, —	1, —	1,041	1,041	1,041	1,041	0,993	0,993	0,993	0,993
	M I_A	1,038	1,218	1,280	1,145	1, —	1, —	1, —	1, —	1,002	1,006	1,056	1,022	0,990	1,103	0,918	
	C I_t	0,995	0,970	0,951	0,814	1, —	1, —	1, —	1, —	1,055	1,080	0,998	0,891	0,963	1,030	0,863	0,851
	C I_P	1,048	1,048	1,048	1,048	1, —	1, —	1, —	1, —	1,041	1,041	1,041	1,041	0,993	0,993	0,993	0,993
	C I_A	1,005	1,031	1,052	1,229	1, —	1, —	1, —	1, —	0,948	0,926	1,002	1,122	1,038	0,971	1,157	1,122
Salerno	M I_t	0,965	0,826	0,825	0,892	1,002	1,006	1,096	1,092	1, —	1, —	1, —	1, —	1,013	1,913	1,151	—
	M I_P	1,006	1,006	1,006	1,006	0,960	0,960	0,960	0,960	1, —	1, —	1, —	1, —	0,954	0,954	0,954	0,954
	M I_A	1,036	1,211	1,212	1,121	0,998	0,994	0,947	0,978	1, —	1, —	1, —	1, —	0,987	1,095	0,869	—
	C I_t	0,943	0,899	0,953	0,913	0,948	0,926	1,002	1,122	1, —	1, —	1, —	1, —	0,914	0,954	0,867	1,010
	C I_P	1,006	1,006	1,006	1,006	0,960	0,960	0,960	0,960	1, —	1, —	1, —	1, —	0,954	0,954	0,954	0,954
	C I_A	1,060	1,112	1,049	1,095	1,055	1,080	0,998	0,891	1, —	1, —	1, —	1, —	1,094	1,048	1,154	0,999
Sassari	M I_t	0,953	0,905	0,717		0,990	1,103	0,978		0,987	1,095	0,869		1, —	1, —	1, —	1, —
	M I_P	1,055	1,055	1,055	1,055	1,007	1,007	1,007	1,007	1,048	1,048	1,048	1,048	1, —	1, —	1, —	1, —
	M I_A	1,049	1,105	1,395		1,055	0,907	1,089		1,013	0,913	1,151		1, —	1, —	1, —	1, —
	C I_t	1,032	0,943	1,100	0,913	1,038	0,971	1,157	1,122	1,094	1,048	1,154	0,999	1, —	1, —	1, —	1, —
	C I_P	1,056	1,055	1,055	1,055	1,007	1,007	1,007	1,007	1,048	1,048	1,048	1,048	1, —	1, —	1, —	1, —
	C I_A	0,969	1,060	0,909	1,095	0,963	1,030	0,863	0,894	0,914	0,954	0,867	1,001	1,—	1, —	1. —	1, —
Messina	M I_t	0,938	0,885	0,876	0,961	0,986	1,090	1,135	1,114	0,971	1,071	1,062	1,077	0,984	0,977	1,222	
	M I_P	0,956	0,956	0,956	0,956	0,913	0,913	0,913	0,913	0,950	0,950	,950	0,950	0,906	0,906	0,906	0,906
	M I_A	1,066	1,130	1,142	1,041	1,014	0,917	0,881	0,898	1,030	0,934	0,942	0,929	1,016	1,024	0,818	
	C I_t	0,852	0,826	0,836	0,827	0,866	0,861	0,890	1,028	0,903	0,918	0,877	0,906	0,825	0,876	0,760	0,960
	C I_P	0,956	0,956	0,956	0,956	0,913	0,913	0,913	0,913	0,950	0,950	0,950	0,950	0,906	0,906	0,906	0,906
	C I_A	1,174	1,211	1,196	1,209	1,155	1,161	1,124	0,973	1,107	1,089	1,140	1,104	1,212	1,142	1,316	1,042
Campobasso	M I_t	1,052	0,859	0,897	0,935	1,092	1,046	1,148	1,070	1,089	1,040	1,088	1,048	1,104	0,949	1,252	
	M I_P	1,023	1,023	1,023	1,023	0,976	0,976	0,976	0,976	1,017	1,017	1,017	1,017	0,970	0,970	0,970	0,970
	M I_A	0,951	1,164	1,115	1,070	0,916	0,956	0,871	0,935	0,918	0,962	0,919	0,954	0,966	1,054	0,829	
	C I_t	0,842	0,845	0,814	0,811	0,846	0,870	0,856	0,995	0,893	0,939	0,855	0,888	0,816	0,896	0,741	0,888
	C I_P	1,023	1,023	1,023	1,023	0,976	0,976	0,976	0,976	1,017	1,017	1,017	1,017	0,970	0,970	0,970	0,970
	C I_A	1,188	1,183	1,229	1,233	1,182	1,149	1,168	1,005	1,120	1,065	1,170	1,126	1,225	1,116	1,350	1,126
Milano	M I_t	1,064	1,017	1,038	1,194	1,105	1,239	1,329	1,368	1,102	1,231	1,258	1,338	1,117	1,124	1,448	
	M I_P	0,970	0,970	0,970	0,970	0,926	0,926	0,926	0,926	0,964	0,964	0,964	0,964	0,920	0,920	0,920	0,920
	M I_A	0,940	0,983	0,963	0,838	0,905	0,807	0,752	0,731	0,839	0,812	0,795	0,747	0,895	0,890	0,691	
	C I_t	1,043	1,039	1,139	1,194	1,049	1,070	1,198	1,467	1,106	1,155	1,195	1,307	1,011	1,101	1,036	1,308
	C I_P	0,970	0,970	0,970	0,970	0,926	0,926	0,926	0,926	0,964	0,964	0,964	0,964	0,920	0,920	0,920	0,920
	C I_A	0,959	0,962	0,878	0,838	0,953	0,935	0,835	0,682	0,904	0,866	0,837	0,765	0,989	0,907	0,965	0,765

Le città scritte nella testata orizzontale stanno a numeratore.

Messina				Campobasso				Milano			
I	II	III	IV	I	II	III	IV	I	II	III	IV
1,066	1,130	1,142	1,041	0,951	1,164	1,115	1,070	0,940	0,983	0,963	0,838
1,046	1.046	1.046	1.046	0,978	0,978	0,978	0,978	1,031	1,031	1,031	1,031
0,938	0.885	0.876	0,961	1.052	0.859	0.897	0.935	1.064	1.017	1.038	1.194
1,174	1,211	1,196	1,209	1,188	1,183	1,229	1,233	0,959	0,962	0,878	0,838
1,046	1,046	1,046	1,046	0,978	0,978	0,978	0,978	1,031	1,031	1,031	1,031
0,052	0,826	0,836	0,827	0,842	0,845	0,814	0,811	1,043	1,039	1,139	1,194
1,014	0,917	0,881	0,898	0,916	0,956	0,871	0,935	0,905	0,807	0,752	0,731
1,096	1,096	1,096	1,096	1,024	1,024	1,024	1,024	1,080	1,080	1,080	1,080
1,986	1,090	1,135	1,114	1.092	1,046	1,148	1,070	1,105	1,289	1,329	1,368
1,155	1,161	1,124	0,973	1,182	1,149	1,168	1,005	0,953	0,935	0,835	0,682
1,096	1,096	1,096	1,096	1,024	1,024	1,024	1,024	1,080	1,080	1,080	1,080
0,866	0,861	0,890	1,028	0,846	0,875	0,856	0,995	1,049	1,070	1,198	1,467
1,030	0,934	0,942	0,929	0,918	0,962	0,919	0,954	0,839	0,812	0,795	0,747
1,053	1,053	1,053	1,053	0,984	0,984	0,984	0,984	1,057	1,037	1,037	1,037
0,971	1,071	1,062	1,077	1,089	1,040	1,088	1,048	1,102	1,231	1,258	1,338
1,107	1.089	1,140	1,104	1,120	1,065	1,170	1,126	0,904	0,866	0,837	0,765
1,053	1,053	1,053	1,053	0,984	0,984	0,984	0,984	1,037	1,037	1,037	1,037
0,903	0,918	0.877	0,906	0,893	0,939	0,855	0,888	1,106	1,115	1,195	1.307
1,016	1.024	0,818		0,906	1.054	0,799		0,895	0,890	0,691	
1,103	1,103	1,103	1,103	1,031	1,031	1,031	1,031	1,087	1,087	1,087	1,087
0,984	0,977	1,222		1,104	0,949	1,252		1,117	1,124	1,448	
1,212	1,142	1,316	1,042	1,225	1,116	1,350	1,126	0,989	0,907	0,965	0,765
1,103	1,103	1,103	1,103	1,031	1,031	1,031	1,031	1,087	1,087	1,087	1,087
0,825	0,876	0,760	0,960	0,816	0,896	0,741	0,888	1,011	1,102	1,036	1,308
1, —	1, —	1, —	1, —	0,892	1,030	0,977	1,028	0,882	0,870	0,844	0,805
1, —	1, —	1, —	1, —	0,935	0,935	0,935	0,935	0,986	0,986	0,986	0,986
1, —	1, —	1, —	1, —	1,121	0,971	1,024	0,973	1,134	1,150	1,185	1,243
1, —	1, —	1, —	1, —	1,012	0,978	1,028	1,020	0,816	0,795	0,734	0,693
1, —	1. —	1, —	1, —	0,935	0,935	0,935	0,935	0,986	0,986	0,986	0,986
1, —	1, —	1, —	1, —	0,988	1,023	0,973	0,980	1,225	1,258	1,362	1,442
1,121	0,971	1,024	0,973	1, —	1, —	1, —	1, —	0,989	0,845	0,864	0,783
1,070	1,070	1,070	1,070	1, —	1, —	1, —	1, —	1,055	1,055	1,055	1,055
0,892	1,030	0,977	1,028	1, —	1, —	1, —	1, —	1,011	1,184	1,157	1,277
0.988	1,023	0,973	0,980	1, —	1, —	1, —	1, —	0,989	0,845	0,864	0,783
1,070	1,070	1,070	1,070	1; —	1, —	1, —	1, —	1,055	1,055	1,055	1,055
1,012	0,978	1,028	1,020	1, —	1, —	1, —	1, —	1,239	1,230	1,398	1,473
1,134	1,150	1,185	1,243	1,011	1,184	1,157	1,277	1, —	1, —	1, —	1, —
1,015	1,015	1,015	1,015	0,948	0,948	0,948	0,948	1, —	1, —	1, —	1, —
0,882	0,870	0,844	0,805	0,989	0,845	0,864	0,783	1, —	1, —	1, —	1, —
1,225	1,258	1,362	1,442	1,239	1,230	1,398	1,473	1, —	1, —	1, —	1, —
1,015	1,015	1,015	1,015	0,948	0,948	0,948	0,948	1, —	1, —	1, —	1, —
0,816	0,795	0,734	0,693	0,807	0,813	0,715	0,679	1, —	1, —	1, —	1,—

M = operai metalmeccanici

C = operai edili

I, II, III, IV indicano rispettivamente:

I = operai specializzati

II = operai quaificati

III = manovali specializz.

IV = manovali comuni

$$ {}_{a,b}I_t = \frac{S_b}{S_a} \frac{\sum_l P_{al} \frac{\sum_{i=1}^{7} P_{il} Q_{il}}{\sum_{i=1}^{7} P_{il}}}{\sum_l P_{bl} \frac{\sum_{i=1}^{7} P_{il} Q_{il}}{\sum_{i=1}^{7} P_{il}}} = $$

$$ = \frac{S_b}{S_a} \, {}_{a,b}I_p $$

$$ {}_{a,b}I_A = \frac{1}{{}_{a,b}I_t} = \frac{S_a}{S_b} \frac{1}{{}_{a,b}I_p} $$

dove:

${}_{a,b}I_p$ = indice dei prezzi della città a rispetto della città b (formula II del Gini *average weights system*)

S_a, S_b = rispett. salari delle città a e b

${}_{a,b}I_t$ = indice dei tempi della città a rispetto alla città b

${}_{a,b}I_A$ = indice del potere di acquisto del salario S_a rispetto a S_b relativamente alle razioni alimentari.

TABELLA 3ª

Tabella degli indici I_t, I_P, I_A, di ogni città rispetto a ciascun altra.

			Bolzano				Lecce				Salerno				Sassari			
			I	II	III	IV	I	II	III	IU	I	II	III	IV	I	II	III	IV
Bolzano	M	I_t	1, —	1, —	1. —	1, —	1,123	1,318	1,385	1,239	1,102	1,288	1,290	1,192	1,094	1,151	1,453	
		I_P	1, —	1, —	1. —	1, —	1,032	1,032	1,032	1,032	1.058	1,058	1,058	1,058	0,988	0,988	0,988	—
		I_A	1, —	1. —	1. —	1, —	0,890	0,759	0,722	0,807	0,907	0,776	0,775	0,839	0,914	0,869	0,688	
	C	I_t	1. —	1. —	1. —	1. —	1,088	1,114	1.137	1,329	1,129	1,183	1,116	1,165	1.010	1,105	0,947	1,142
		I_P	1. —	1. —	1. —	1. —	1,032	1,032	1,032	1,032	1,058	1,058	1,058	1,058	0,988	0,988	0,988	0,988
		I_A	1, —	1. —	1, —	1. —	0,919	0,897	0,879	0,752	0,886	0,845	0,896	0,858	0,990	0,905	1,055	0,876
Lecce	M	I_t	0,890	0,759	0,722	0,807	1, —	1, —	1, —	1, —	0,981	0,978	0,934	0,662	0,973	0,874	1,049	
		I_P	0,969	0,969	0,969	0,969	1, —	1, —	1, —	1, —	1,024	1,024	1,024	1,024	0,957	0,957	0,957	—
		I_A	1,123	1,318	1,385	1,239	1. —	1, —	1. —	1. —	1.019	1,023	1,074	1,039	1,027	1,145	0,953	
	C	I_t	0,919	0,897	0,879	0,752	1, —	1, —	1, —	1, —	1,038	1,062	0,982	0,877	0,928	0,992	0,833	0,859
		I_P	0,969	0,969	0,969	0,969	1, —	1, —	1, —	1, —	1,024	1,024	1,024	1,024	0,957	0,957	0,957	0,957
		I_A	1,088	1,114	1,137	1,329	1, —	1, —	1, —	1, —	0,964	0,942	1,019	1,141	1,077	1,008	1,200	1,164
Salerno	M	I_t	0,907	0,776	0,775	0,839	1,019	1,023	1,074	1,039	1, —	1. —	1, —	1. —	0,992	0,894	1,126	
		I_P	0,945	0,945	0,945	0,945	0,976	0,976	0,678	0,976	1, —	1, —	1. —	1. —	0,934	0,934	0,934	—
		I_A	1,102	1 288	1,290	1,192	0,981	0,978	0,934	0,962	1. —	1, —	1, —	1. —	1,008	1,119	0,888	
	C	I_t	0,886	0,845	0,896	0,858	0,964	0,942	1,019	1,141	1, —	1, —	1. —	1. —	0,895	0,934	0,849	0,980
		I_P	0,945	0,945	0,945	0,945	0,976	0,976	0,976	0,976	1, —	1. —	1, —	1. —	0,934	0,934	0,934	0,934
		I_A	1,129	1,183	1,116	1,165	1,038	1,062	0,982	0,877	1, —	1. —	1, —	1, —	1,118	1,071	1,178	1,021
Sassari	M	I_t	0,914	0,869	0,688		1,027	1,145	0,953		1,008	1,119	0,888		1. —	1. —	1, —	1, —
		I_P	1.012	1,012	1,012	—	1,045	1.045	1.045	—	1,071	1,071	1,071	—	1. —	1,	1. —	1, —
		I_A	1.094	1,151	1,453		0,973	0,874	1,049		0,992	0,894	1,126		1. —	1. —	1, —	1. —
	C	I_t	0,990	0,905	1,055	0,876	1,077	1,008	1,200	1,164	1,118	1,071	1,178	1,021	1. —	1.	1. —	1,
		I_P	1,012	1,012	1,012	1,012	1,045	1.045	1,045	1,045	1,071	1,071	1,071	1,071	1. —	1. —	1,	1.
		I_A	1,010	1,105	0,947	1,142	0,928	0,992	0,833	0,859	0,895	0,934	0,849	0,980	1. —	1.	1. —	1. —
Messina	M	I_t	0,878	0,828	0,820	0,899	0,986	1,091	1,114	1,114	0,968	1,066	1,057	1,072	0,960	0,953	1,192	
		I_P	0,894	0,894	0,894	0,894	0,923	0,923	0,923	0,923	0,946	0,946	0,946	0,946	0,884	0,884	0,884	—
		I_A	1,139	1,208	1,220	1,112	1,014	0,917	0,898	0,898	1,034	0,938	0,946	0,932	1,042	1,049	0,839	
	C	I_t	0,797	0,773	0,783	0,774	0,867	0,861	0,890	1,029	0,899	0,914	0,874	0,902	0,805	0,854	0,741	0,884
		I_P	0,894	0,894	0,894	0,894	0,923	0,923	0,923	0,923	0,946	0,946	0,946	0,946	0,884	0,884	0,884	0,884
		I_A	1,255	1,294	1,278	1,292	1,154	1,161	1,124	0,972	1,111	1,094	1,145	1,109	1,243	1,171	1,349	1,132
Campobasso	M	I_t	0,975	0,797	0,832	0,867	1,096	1,050	1,153	1,075	1,075	1,026	1,108	1,034	1,067	0,917	1,209	
		I_P	0,949	0,949	0,949	0,949	0,980	0,980	0,980	0,980	1,004	1,004	1,004	1,004	0,937	0,937	0,937	—
		I_A	1,025	1,255	1,201	1,153	0,912	0,953	0,869	0,931	0,930	0,974	0,903	0,967	0,938	1,091	0,826	
	C	I_t	0,781	0,783	0,755	0,752	0,849	0,873	0,859	0,999	0,881	0,927	0,843	0,876	0,822	0,866	0,716	0,858
		I_P	0,949	0,949	0,949	0,949	0,980	0,980	0,980	0,980	1,004	1,004	1,004	1,004	0,937	0,937	0,937	0,937
		I_A	1,281	1. 76	1,324	1,330	1,177	1,145	1,164	1,001	1,135	1,079	1,186	1,142	1,216	1,155	1,397	1,165
Milano	M	I_t	1,007	0,963	0,983	1,131	1,193	1,269	1,361	1,401	1,111	1,241	1,268	1,349	1,102	1,109	1,429	
		I_P	0,919	0,919	0,919	0,919	0,949	0,949	0,949	0,949	0,972	0,972	0,972	0,972	0,908	0,908	0,908	0,908
		I_A	0,993	1,038	1,017	0,884	0,838	0,788	0,734	0,714	0,900	0,806	0,789	0,741	0,908	0,902	0,700	
	C	I_t	0,988	0,984	1,079	1,131	1,075	1,096	1,227	1,503	1,115	1,164	1,205	1,318	0,998	1,087	1,022	1,291
		I_P	0,919	0,919	0,919	0,919	0,949	0,949	0,949	0,949	0,972	0,972	0,972	0,972	0,908	0,908	0,908	0,908
		I_A	1,012	1,016	0,927	0,884	0,930	0,912	0,815	0,665	0,897	0,869	0,830	0,759	1,002	0,920	0,978	0,775

e città scritte nella testata orizzontale stanno a numeratore.

Messina				Campobasso				Milano			
I	II	III	IV	I	II	III	IV	I	II	III	IV
,139	1,208	1,220	1,112	1,025	1,255	1,201	1,153	0,993	1,038	1,017	0,884
,118	1,118	1,118	1,118	1,054	1,054	1,054	1,054	1,088	1,088	1,088	1,088
,878	0,828	0,820	0,899	0,975	0,797	0,832	0,867	1,007	0,963	0,983	1,131
,255	1,294	1,278	1,292	1,281	1,276	1,324	1,330	1,012	1,016	0,927	0,884
,118	1,118	1,118	1,118	1,054	1,054	1,054	1,054	1,088	1,088	1,088	1,088
,797	0,773	0,783	0,774	0,781	0,783	0,755	0,752	0,988	0,984	1,079	1,131
,014	0,917	0,898	0,898	0,912	0,953	0,869	0,931	0,838	0,788	0,734	0,714
,083	1,083	1,083	1,083	1,021	1,021	1,021	1,021	1,054	1,054	1,054	1,054
,986	1,091	1,114	1,114	1,096	1,050	1,153	1,075	1,193	1,269	1,361	1,401
,154	1,161	1,124	0,972	1,177	1,145	1,164	1,001	0,930	0,912	0,815	0,665
,083	1,083	1,083	1,083	1,021	1,021	1,021	1,021	1,054	1,054	1,054	1,054
,867	0,861	0,890	1,029	0,849	0,873	0,859	0,999	1,075	1,096	1,227	1,503
,034	0,938	0,946	0,932	0,930	0,974	0,903	0,967	0,900	0,806	0,789	0,741
,106	1,106	1,106	1,106	0,996	0,996	0,996	0,996	1,029	1,029	1,029	1,029
,968	1,066	1,057	1,072	1,075	1,026	1,108	1,034	1,111	1,241	1,268	1,349
,111	1,094	1,145	1,109	1,135	1,079	1,186	1,142	0,897	0,859	0,830	0,759
,106	1,106	1,106	1,106	0,996	0,996	0,996	0,996	1,029	1,029	1,029	1,029
,899	0,914	0,874	0,902	0,881	0,927	0,843	0,876	1,115	1,164	1,205	1,318
,042	1,049	0,839		0,938	1,091	0,827		0,908	0,902	0,700	
,132	1,132	1,132	—	1,077	1,077	1,077	—	1,102	1,102	1,102	—
,960	0,953	1,192		1,067	0,917	1,209		1,102	1,109	1,429	
,243	1,171	1,349	1,132	1,216	1,155	1,397	1,165	1, —	0,920	0,978	0,775
,132	1,132	1,132	1,132	1,077	1,077	1,077	1,077	1,102	1,102	1,102	1,102
,805	0,854	0,741	0,884	0,822	0,866	0,716	0,858	0,998	1,088	1,022	1,291
, —	1, —	1, —	1, —	0,900	1,039	0,985	1,037	0,871	0,859	0,834	0,795
, —	1, —	1, —	1, —	0,943	0,943	0,943	0,943	0,974	0,974	0,974	0,974
, —	1, —	1, —	1, —	1,112	0,963	1,015	0,964	1,148	1,164	1,199	1,258
, —	1, —	1, —	1, —	1,020	0,986	1,036	1,029	0,807	0,785	0,725	0,685
, —	1, —	1, —	1, —	0,943	0,943	0,943	0,943	0,974	0,974	0,974	0,974
, —	1, —	1, —	1, —	0,980	1,014	0,965	0,971	1,240	1,273	1,379	1,461
,115	0,963	1,015	0,964	1,	1, —	1, —	1, —	0,968	0,827	0,847	0,767
,061	1,061	1,061	1,061	1, —	1, —	1, —	1, —	1,033	1,033	1,033	1,033
,900	1,039	0,985	1,037	1, —	1, —	1, —	1, —	1,033	1,209	1,181	1,304
,980	1,014	0,965	0,971	1, —	1, —	1, —	1, —	0,790	0,796	0,700	0,665
,061	1,061	1,061	1,061	1, —	1, —	1, —	1, —	1,033	1,033	1,033	1,033
,020	0,986	1,036	1,029	1, —	1, —	1, —	1, —	1,265	1,256	1,429	1,504
,148	1,164	1,199	1,258	1,033	1,209	1,181	1,304	1, —	1, —	1, —	1, —
,027	1,027	1,027	1,027	0,968	0,968	0,968	0,968	1, —	1, —	1, —	1, —
,871	0,859	0,834	0,795	0,968	0,827	0,847	0,767	1, —	1, —	1, —	1, —
,240	1,273	1,379	1,461	1,265	1,256	1,428	1,504	1, —	1, —	1, —	1, —
,027	1,027	1,027	1,027	0,968	0,968	0,968	0,968	1, —	1, —	1, —	1, —
,807	0,785	0,725	0,685	0,790	0,796	0,700	0,665	1, —	1, —	1, —	1, —

M = operai metalmeccanici
C = operai edili
I, II, III, IV indicano le classi rispettiv.

I = operai specializzati
II = operai qualificati
III = manovali specializzati
IV = manovali comuni

TABELLA 4ª

Tabella degli indici I_t, I_P, I_A, di ogni città rispetto a ciascun altra

			Bolzano				Lecce				Salerno				Sassari			
			I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
Bolzano . .	M	I_t					1,111	1,101	1,125	1,314	1,086	1,269	1,271	1,174	1,092	1,149	1,451	
		I_P	1, —	1, —	1, —	1, —	1,021	1,021	1,021	1,021	1,042	1,042	1,042	1,042	0,986	0,986	0,986	
		I_A					0,900	0,768	0,731	0,816	0,921	0,788	0,787	0,852	0,916	0,870	0,689	
	C	I_t					1,075	1,101	1,125	1,314	1,111	1,166	1,099	1,147	1,008	1,104	0,946	1,140
		I_P	1, —	1, —	1, —	1, —	1,021	1,021	1,021	1,021	1,042	1,042	1,042	1,042	0,986	0,986	0,986	0,986
		I_A					0,930	0,908	0,889	0,761	0,900	0,858	0,910	0,872	0,992	0,906	1,057	0,877
Lecce . . .	M	I_t	0,900	0,768	0,731	0,816					0,978	0,975	0,928	0,959	0,983	0,883	1,059	
		I_P	0,980	0,980	0,980	0,980	1, —	1, —	1, —	1, —	1,021	1,021	1,021	1,021	0,966	0,966	0,966	0,966
		I_A	1,111	1,308	1,368	1,225					1,023	1,026	1,078	1,043	1,017	1,133	0,944	
	C	I_t	0,930	0,908	0,889	0,761					1,034	1,058	0,978	0,873	0,937	1,002	0,841	0,867
		I_P	0,980	0,980	0,980	0,980	1, —	1, —	1, —	1, —	1,021	1,021	1,021	1,021	0,966	0,966	0,966	0,966
		I_A	1,075	1,101	1,125	1,314					0,967	0,945	1,022	1,145	1,067	0,998	1,189	1,153
Salerno. .	M	I_t	0,921	0,788	0,787	0,852	1,023	1,026	1,078	1,043					1,006	0,906	1,143	
		I_P	0,960	0,960	0,960	0,960	0,980	0,980	0,980	0,980	1, —	1, —	1, —	1, —	0,947	0,947	0,947	0,947
		I_A	1,086	1,269	1,271	1,174	0,978	0,975	0,928	0,959					0,994	1,104	0,975	
	C	I_t	0,900	0,858	0,910	0,872	0,967	0,945	1,022	1,145					0,907	0,947	0,861	0,993
		I_P	0,960	0,960	0,960	0,960	0,980	0,980	0,980	0,980	1, —	1, —	1, —	1, —	0,947	0,947	0,947	0,947
		I_A	1,111	1,166	1,099	1,147	1,034	1,058	0,978	0,873					1,102	1,056	1,162	1,007
Sassari. .	M	I_t	0,916	0,870	0,689		1,017	1,133	0,944		0,994	1,104	0,875					
		I_P	1,014	1,014	1,014	—	1,035	1,035	1,035	—	1,056	1,056	1,056	—	1, —	1, —	1, —	1, —
		I_A	1,092	1,149	1,451		0,983	0,883	1,059		1,006	0,906	1,143					
	C	I_t	0,992	0,906	1,057	0,877	1,067	0,998	1,189	1,153	1,102	1,056	1,162	1,007				
		I_P	1,014	1,014	1,014	1,014	1,035	1,035	1,035	1,035	1,056	1,056	1,056	1,056	1, —	1, —	1, —	1, —
		I_A	1,008	1,104	0,946	1,140	0,937	1,002	0,841	0,867	0,907	0,947	0,861	0,993				
Messina . .	M	I_t	0,886	0,835	0,828	0,908	0,984	1,089	1,133	1,112	0,962	1,060	1,051	1,066	0,968	0,961	1,201	
		I_P	0,903	0,903	0,903	0,903	0,922	0,922	0,922	0,922	0,940	0,940	0,940	0,940	0,891	0,891	0,891	0,891
		I_A	1,129	1,198	1,208	1,101	1,016	0,918	0,883	0,899	1,039	0,943	0,951	0,938	1,033	1,041	0,933	
	C	I_t	0,804	0,780	0,790	0,781	0,865	0,859	0,888	1,027	0,894	0,909	0,869	0,897	0,811	0,861	0,748	0,891
		I_P	0,903	0,903	0,903	0,903	0,922	0,922	0,922	0,922	0,940	0,940	0,940	0,940	0,891	0,891	0,891	0,891
		I_A	1,244	1,282	1,266	1,280	1,156	1,164	1,126	0,974	1,119	1,100	1,151	1,115	1,233	1,161	1,337	1,122
Campobasso	M	I_t	1,002	0,818	0,855	0,891	1,113	1,066	1,170	1,091	1,087	1,038	1,085	1,045	1,101	0,946	1,248	
		I_P	0,975	0,975	0,975	0,975	0,995	0,995	0,995	0,995	1,015	1,015	1,015	1,015	0,967	0,967	0,967	0,967
		I_A	0,998	1,222	1,170	1,122	0,898	0,938	0,855	0,917	0,920	0,963	0,922	0,957	0,908	1,057	0,801	
	C	I_t	0,802	0,805	0,776	0,772	0,862	0,886	0,872	1,014	0,891	0,937	0,853	0,886	0,814	0,893	0,738	0,886
		I_P	0,975	0,975	0,975	0,975	0,995	0,995	0,995	0,995	1,015	1,015	1,015	1,015	0,967	0,967	0,967	0,967
		I_A	1,247	1,242	1,289	1,295	1,160	1,129	1,147	0,986	1,122	1,067	1,172	1,129	1,229	1,120	1,355	1,129
Milano. . .	M	I_t	1,036	0,991	1,011	1,163	1,151	1,291	1,385	1,425	1,124	1,257	1,284	1,365	1,132	1,140	1,468	
		I_P	0,945	0,945	0,945	0,945	0,965	0,965	0,965	0,965	0,984	0,984	0,984	0,984	0,933	0,933	0,933	0,933
		I_A	0,965	1,009	0,989	0,860	0,869	0,775	0,722	0,702	0,890	0,796	0,779	0,733	0,883	0,877	0,681	
	C	I_t	1,016	1,012	1,109	1,163	1,283	1,115	1,248	1,528	1,129	1,179	1,219	1,334	1,025	1,118	1,051	1,327
		I_P	0,945	0,945	0,945	0,945	0,965	0,965	0,965	0,965	0,984	0,984	0,984	0,984	0,933	0,933	0,933	0,933
		I_A	0,984	0,988	0,902	0,860	0,779	0,897	0,801	0,654	0,886	0,848	0,820	0,859	0,976	0,894	0,951	0,754

Le città scritte nella testata orizzontale stanno a numeratore.

Messina				Campobasso				Milano			
I	II	III	IV	I	II	III	IV	I	II	III	IV
1,129	1,198	1,208	1,101	0,998	1,222	1,170	1,122	0,965	0,991	1,011	0,859
1,108	1,108	1,108	1,108	1,026	1,026	1,026	1,026	1,058	1,058	1,058	1,058
0,886	0,835	0,828	0,908	1,002	0,818	0,855	0,891	1,036	0,991	1,011	1,163
1,244	1,282	1,266	1,280	1,247	1,242	1,289	1,295	0,984	0,988	0,902	0,860
1,108	1,108	1,108	1,108	1,026	1,026	1,026	1,026	1,058	1,058	1,058	1,058
0,804	0,780	0,790	0,781	0,802	0,805	0,776	0,772	1,016	1,012	1,109	1,163
1,016	0,918	0,883	0,899	0,898	0,938	0,855	0,917	0,869	0,775	0,722	0,702
1,085	1,085	1,085	1,085	1,005	1,005	1,005	1,005	1,037	1,037	1,037	1,037
0,984	1,089	1,133	1,112	1,113	1,066	1,170	1,091	1,151	1,294	1,385	1,425
1,156	1,164	1,126	0,974	1,160	1,129	1,147	0,986	0,779	0,897	0,801	0,654
1,085	1,085	1,085	1,085	1,005	1,005	1,005	1,005	1,037	1,037	1,037	1,037
0,865	0,859	0,888	1,027	0,862	0,886	0,872	1,014	1,283	1,115	1,248	1,528
1,039	0,943	0,951	0,938	0,920	0,963	0,922	0,957	0,890	0,796	0,779	0,733
1,063	1,063	1,063	1,063	0,986	0,986	0,986	0,986	1,016	1,016	1,016	1,016
0,962	1,060	1,051	1,066	1,087	1,038	1,085	1,045	1,124	1,257	1,284	1,365
1,119	1,100	1,151	1,115	1,122	1,067	1,172	1,129	0,886	0,848	0,820	0,759
1,063	1,063	1,063	1,063	0,986	0,986	0,986	0,986	1,016	1,016	1,016	1,016
0,894	0,909	0,869	0,897	0,891	0,937	0,853	0,886	1,129	1,179	1,219	1,334
1,033	1,041	0,833		0,908	1,057	0,801		0,883	0,877	0,681	
1,122	1,122	1,122	1,122	1,034	1,034	1,034	1,034	1,072	1,072	1,072	—
0,968	0,961	1,201		1,101	0,946	1,248		1,132	1,140	1,468	
1,233	1,161	1,337	1,122	1,229	1,120	1,355	1,129	0,976	0,894	0,951	0,754
1,122	1,122	1,122	1,122	1,034	1,034	1,034	1,034	1,072	1,072	1,072	1,072
0,811	0,861	0,748	0,891	0,814	0,893	0,738	0,886	1,025	1,118	1,051	1,327
				0,884	1,021	0,968	1,019	0,855	0,844	0,819	0,781
1, —	1, —	1, —	1, —	0,927	0,927	0,927	0,927	0,956	0,956	0,956	0,956
				1,131	0,979	1,033	0,981	1,169	1,185	1,221	1,281
				1,003	0,969	1,018	1,012	0,792	0,771	0,712	0,672
1, —	1, —	1, —	1, —	0,927	0,927	0,927	0,927	0,956	0,956	0,956	0,956
				0,997	1,032	0,982	0,988	1,263	1,297	1,404	1,488
1,131	0,979	1,033	0,981					0,967	0,826	0,845	0,766
1,079	1,079	1,079	1,079	1, —	1, —	1, —	1, —	1,031	1,031	1,031	1,031
0,884	1,021	0,968	1,019					1,034	1,211	1,183	1,306
0,997	1,032	0,982	0,988					0,789	0,795	0,699	0,664
1,079	1,079	1,079	1,079	1, —	1, —	1, —	1, —	1,031	1,031	1,031	1,031
1,003	0,969	1,018	1,012					1,267	1,258	1,431	1,506
1,169	1,185	1,221	1,281	1,034	1,211	1,183	1,306				
1,046	1,046	1,046	1,046	0,970	0,970	0,970	0,970	1, —	1, —	1, —	1, —
0,855	0,844	0,819	0,781	0,967	0,826	0,845	0,766				
1,263	1,297	1,404	1,488	1,267	1,258	1,431	1,506				
1,046	1,046	1,046	1,046	0,970	0,970	0,970	0,970	1, —	1, —	1, —	1, —
0,792	0,771	0,712	0,672	0,789	0,795	0,699	0,664				

M = operai metalmeccanici
C = operai edili
I, II, III e IV indicano rispettivamente :
I = operai specializzati
II = operai qualificati
III = manovali specializzati
IV = manovali comuni

TABELLA 5ª
Dati secondo la

		Bolzano				Lecce				Salerno				Sassari			
		I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
Bolzano	M I_t					1,127	1,323	1,382	1,244	1,052	1,229	1,230	1,138	1,122	1,182	1,493	
	M I_P	1, —	1, —	1, —	1, —	1,036	1,036	1,036	1,036	1,009	1,009	1,009	1,009	1,015	1,015	1,015	1,015
	M I_A					0,887	0,756	0,720	0,804	0,951	0,814	0,813	0,879	0,891	0,846	0,670	
	C I_t					1,092	1,119	1,142	1,334	1,078	1,129	1,065	1,111	1,037	1,135	0,973	1,172
	C I_P	1, —	1, —	1, —	1, —	1,036	1,036	1,036	1,036	1,009	1,009	1,009	1,009	1,015	1,015	1,015	1,015
	C I_A					0,916	0,894	0,876	0,750	0,928	0,886	0,939	0,900	0,964	0,881	1,028	0,853
Lecce	M I_t	0,887	0,756	0,720	0,804					1,006	1,003	0,955	0,987	0,970	0,870	1,046	—
	M I_P	0,965	0,965	0,965	0,965	1, —	1, —	1, —	1, —	1,051	1,051	1,051	1,051	0,953	0,953	0,953	0,953
	M I_A	1,127	1,323	1,389	1,244					0,994	0,997	1,047	1,013	1,031	1,149	0,956	—
	C I_t	0,916	0,894	0,876	0,750					1,064	1,089	1,007	0,899	0,925	0,988	0,830	0,856
	C I_P	0,965	0,965	0,965	0,965	1, —	1, —	1, —	1, —	1,051	1,051	1,051	1,051	0,953	0,953	0,953	0,953
	C I_A	1,092	1,119	1,142	1,333					0,940	0,918	0,993	1,112	1,081	1,012	1,205	1,168
Salerno	M I_t	0,951	0,814	0,813	0,879	0,994	0,997	1,047	1,013					0,999	0,900	1,135	
	M I_P	0,991	0,991	0,991	0,991	0,952	0,952	0,952	0,952	1, —	1, —	1, —	1, —	0,941	0,941	0,941	0,941
	M I_A	1,052	1,229	1,230	1,138	1,006	1,003	0,955	0,987					1,001	1,111	0,881	—
	C I_t	0,928	0,886	0,939	0,900	0,940	0,918	0,993	1,112					0,901	0,941	0,855	0,987
	C I_P	0,991	0,991	0,991	0,991	0,952	0,952	0,952	0,952	1, —	1, —	1, —	1, —	0,941	0,941	0,941	0,941
	C I_A	1,078	1,129	1,065	1,111	1,064	1,089	1,007	0,899					1,110	1,063	1,170	1,013
Sassari	M I_t	0,891	0,846	0,670	—	1,031	1,149	0,956	—	1,001	1,111	0,881	—				
	M I_P	0,986	0,986	0,986	0,986	1,049	1,049	1,049	1,049	1,063	1,063	1,063	1,063	1, —	1, —	1, —	1, —
	M I_A	1,122	1,182	1,493		0,970	0,870	1,046		0,999	0,900	1,135					
	C I_t	0,964	0,881	1,028	0,853	1,081	1,012	1,205	1,168	1,110	1,063	1,170	1,013				
	C I_P	0,986	0,986	0,986	0,986	1,049	1,049	1,049	1,049	1,063	1,063	1,063	1,063	1, —	1, —	1, —	1, —
	C I_A	1,037	1,135	0,973	1,172	0,925	0,988	0,830	0,856	0,901	0,941	0,855	0,987				
Messina	M I_t	0,856	0,807	0,799	0,877	1,064	1,176	1,224	1,202	0,963	1,062	1,053	1,068	0,931	0,924	1,156	—
	M I_P	0,872	0,872	0,872	0,872	0,996	0,996	0,996	0,996	0,942	0,942	0,942	0,942	0,857	0,857	0,857	0,857
	M I_A	1,168	1,239	1,252	1,140	0,940	0,850	0,817	0,832	1,038	0,942	0,950	0,936	1,074	1,082	0,865	
	C I_t	0,891	0,753	0,763	0,755	0,935	0,929	0,960	1,109	0,895	0,910	0,870	0,898	0,780	0,828	0,719	0,857
	C I_P	0,872	0,872	0,872	0,872	0,996	0,996	0,996	0,996	0,942	0,942	0,942	0,942	0,857	0,857	0,857	0,857
	C I_A	1,122	1,328	1,311	1,325	1,070	1,076	1,042	0,902	1,117	1,099	1,149	1,114	1,282	1,208	1,391	1,167
Campobasso	M I_t	1,007	0,823	0,860	0,896	1,058	1,014	1,113	1,038	1,082	1,033	1,080	1,041	1,122	0,965	1,272	—
	M I_P	0,980	0,980	0,980	0,980	0,946	0,946	0,946	0,946	1,010	1,010	1,010	1,010	0,986	0,986	0,986	0,986
	M I_A	0,993	1,215	1,163	1,116	0,945	0,986	0,898	0,963	0,924	0,968	0,926	0,961	0,891	1,036	0,786	
	C I_t	0,806	0,809	0,780	0,776	0,820	0,843	0,829	0,965	0,887	0,933	0,849	0,882	0,829	0,911	0,753	0,903
	C I_P	0,980	0,980	0,980	0,980	0,946	0,946	0,946	0,946	1,010	1,010	1,010	1,010	0,986	0,986	0,986	0,986
	C I_A	1,241	1,236	1,282	1,289	1,220	1,186	1,206	1,036	1,127	1,072	1,178	1,134	1,206	1,098	1,328	1,107
Milano	M I_t	1,079	1,032	1,053	1,211	1,122	1,258	1,349	1,389	1,122	1,254	1,281	1,363	1,137	1,145	1,475	—
	M I_P	0,984	0,984	0,984	0,984	0,940	0,940	0,940	0,940	0,982	0,982	0,982	0,982	0,937	0,937	0,937	0,937
	M I_A	0,927	0,969	0,945	0,826	0,891	0,795	0,741	0,720	0,891	0,797	0,781	0,734	0,880	0,873	0,678	
	C I_t	1,058	1,054	1,062	1,211	1,065	1,087	1,216	1,489	1,127	1,177	1,217	1,332	1,030	1,123	1,055	1,333
	C I_P	0,984	0,984	0,984	0,984	0,940	0,940	0,940	0,940	0,982	0,982	0,982	0,982	0,937	0,937	0,937	0,937
	C I_A	0,945	0,949	0,942	0,826	0,939	0,920	0,822	0,672	0,887	0,850	0,822	0,751	0,971	0,890	0,948	0,750

Formula ideale »

Messina				Campobasso				Milano			
I	II	III	IV	I	II	III	IV	I	IJ	III	IV
1,168	1,239	1,252	1,140	0,993	1,215	1,163	1,116	0,927	0,969	0,945	0,826
1,147	1,147	1,147	1,147	1,021	1,021	1,021	1,021	1,061	1,016	1,016	1,016
0,856	0,807	0,799	0,877	1,007	0,823	0,860	0,896	1,079	1,032	1,053	1,211
1,122	1,328	1,311	1,325	1,241	1,236	1,282	1,289	0,945	0,949	0,942	0,826
1,147	1,147	1,147	1,147	1,021	1,021	1,021	1,021	1,016	1,016	1,016	1,016
0,894	0,753	0,769	0,755	0,806	0,809	0,780	0,776	1,058	1,054	1,062	1,211
0,940	0,850	0,817	0,832	0,945	0,986	0,898	0,963	0,891	0,795	0,741	0,802
1,004	1,004	1,004	1,004	1,058	1,058	1,058	1,058	1,064	1,064	1,064	1,064
1,064	1,176	1,224	1,202	1,058	1,014	1,113	1,038	1,122	1,258	1,349	1,389
1,070	1,076	1,042	0,902	1,220	1,186	1,206	1,036	0,939	0,920	0,822	0,672
1,004	1,004	1,004	1,004	1,058	1,058	1,058	1,058	1,064	1,064	1,064	1,064
0,935	0,929	0,960	1,109	0,820	0,843	0,829	0,965	1,065	1,087	1,216	1,489
1,038	0,942	0,950	0,936	0,924	0,968	0,926	0,961	0,891	0,797	0,781	0,734
1,061	1,061	1,061	1,061	0,990	0,990	0,990	0,990	1,018	1,018	1,018	1,018
0,963	1,062	1,053	1,068	1,082	1,033	1,080	1,041	1,122	1,254	1,281	1,363
1,117	1,099	1,149	1,114	1,127	1,072	1,178	1,134	0,887	0,850	0,822	0,751
1,061	1,061	1,061	1,061	0,990	0,990	0,990	0,990	1,018	1,018	1,018	1,018
0,895	0,910	0,870	0,898	0,887	0,933	0,849	0,882	1,127	1,177	1,217	1,332
1,074	1,082	0,865		0,891	1,036	0,786	—	0,880	0,873	0,678	—
1,167	1,167	1,167	1,167	1,015	1,015	1,015	1,015	1,067	1,067	1,016	1,067
0,931	0,924	1,156	—	1,122	0,965	1,272	—	1,137	1,145	1,475	—
1,282	1,208	1,391	1,167	1,206	1,098	1,328	1,107	0,971	0,890	0,948	0,750
1,167	1,167	1,167	1,167	1,015	1,015	1,015	1,015	1,067	1,067	1,067	1,067
0,780	0,828	0,719	0,857	0,829	0,911	0,753	0,903	1,030	1,123	1,055	1,333
				0,886	1,022	0,969	1,020	0,850	0,838	0,814	0,775
1, —	1, —	1, —	1, —	0,927	0,927	0,927	0,927	0,949	0,949	0,949	0,949
				1,129	0,978	1,032	0,980	1,177	1,193	1,229	1,290
				1,004	0,971	1,019	1,013	0,787	0,766	0,708	0,668
1, —	1, —	1, —	1, —	0,927	0,927	0,927	0,927	0,949	0,949	0,949	0,949
				0,996	1,030	0,981	0,987	1,271	1,305	1,413	1,498
1,129	0,978	1,032	0,980					0,992	0,847	0,867	0,786
1,078	1,078	1,078	1,078	1, —	1, —	1, —	1, —	1,058	1,058	1,058	1,058
0,886	1,022	0,969	1,020					1,008	1,180	1,153	1,273
0,996	1,030	0,981	0,987					0,810	0,816	0,717	0,681
1,078	1,078	1,078	1,078	1, —	1, —	1, —	1, —	1,058	1,058	1,058	1,058
1,004	0,971	1,019	1,013					1,235	1,226	1,394	1,468
1,177	1,193	1,229	1,290	1,008	1,180	1,153	1,273				
1,053	1,053	1,053	1,053	0,945	0,945	0,945	0,945	1, —	1, —	1, —	1, —
0,850	0,838	0,814	0,775	0,992	0,847	0,867	0,786				
1,271	1,305	1,413	1,498	1,235	1,226	1,394	1,468				
1,053	1,053	1,053	1,053	0,945	0,945	0,945	0,945	1, —	1, —	1, —	1, —
0,787	0,766	0,708	0,668	0,810	0,816	0,717	0,681				

TABELLA 5ª BIS

Numeri indici $\frac{S_a}{S_b}$ *dei salari orari di*

		Bolzano				Lecce				Salerno				Sassari			
		I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
Bolzano	M	1, —	1, —	1, —	1, —	0,919	0,784	0,746	0,833	0,959	0,821	0,820	0,887	0,903	0,858	0,680	—
	C	1, —	1, —	1, —	1, —	0,949	0,926	0,908	0,777	0,937	0,894	0,947	0,908	0,978	0,894	1,043	0,865
Lecce	M	1,088	1,276	1,341	1,200	1, —	1, —	1, —	1, —	1,044	1,047	1,100	1,065	0,983	1,095	0,911	—
	C	1,054	1,079	1,101	1,287	1, —	1, —	1, —	1, —	0,987	0,965	1,043	1,168	1,031	0,965	1,149	1,114
Salerno	M	1,422	1,218	1,219	1,127	0,958	0,955	0,909	0,939	1, —	1, —	1, —	1, —	0,942	1,045	0,829	—
	C	1,067	1,119	1,056	1,102	1,013	1,037	0,958	0,856	1, —	1, —	1, —	1, —	1,044	1, —	1,101	0,953
Sassari	M	1,107	1,165	1,471	—	1,017	0,913	1,097	—	1,062	0,957	1,206	—	1, —	1, —	1, —	1, —
	C	1,022	1,119	0,959	1,156	0,970	1,037	0,871	0,898	0,958	1, =	0,909	1,049	1, —	1, —	1, —	1, —
Messina	M	1,019	1,081	1,091	0,995	0,937	0,847	0,813	0,829	0,978	0,887	0,895	0,882	0,921	0,927	0,742	—
	C	1,123	1,158	1,143	1,156	1,065	1,073	1,038	0,898	1,052	1,035	1,083	1,049	1,098	1,035	1,192	1, —
Campobas.	M	0,973	1,191	1,140	1,094	0,894	0,933	0,850	0,912	0,933	0,978	0,935	0,971	0,879	1,022	0,775	—
	C	1,215	1,211	1,256	1,262	1,153	1,122	1,140	0,981	1,139	1,083	1,190	1,146	1,189	1,083	1,310	1,092
Milano	M	0,912	0,954	0,934	0,812	0,838	0,747	0,697	0,677	0,875	0,783	0,766	0,721	0,824	0,819	0,635	—
	C	0,930	0,934	0,852	0,812	0,883	0,865	0,773	0,631	0,871	0,835	0,807	0,737	0,910	0,835	0,888	0,703

ni città rispetto a ciascuna altra.

Messina				Campobasso				Milano			
I	II	III	IV	I	II	III	IV	I	II	III	IV
,981	0,925	0,917	1,005	1,028	0,840	0,877	0,914	1,097	1,049	1,070	1,231
,891	0,864	0,875	0,865	0,823	0,826	0,796	0,792	1,075	1,071	1,174	1,231
,068	1,181	1,229	1,206	1,119	1,071	1,176	1,097	1,193	1,338	1,435	1,477
,939	0,932	0,964	1,114	0,867	0,891	0,878	1,020	1,074	1,096	1,227	1,503
,023	1,127	1,118	1,133	1,071	1,026	1,108	1,034	1,143	1,277	1,305	1,388
,951	0,966	0,924	0,953	0,878	0,924	0,840	0,873	1,148	1,198	1,240	1,356
,086	1,078	1,348	—	1,133	0,978	0,937	—	1,214	1,222	1,574	—
,921	0,966	0,839	1, —	0,841	0,924	0,763	0,916	1,099	1,198	1,126	1,142
, —	1, —	1, —	1, —	1,048	0,907	0,957	0,909	1,117	1,133	1,167	1,225
, —	1, —	1, —	1, —	0,924	0,956	0,910	0,916	1,207	1,240	1,342	1,422
,955	1,102	1,045	1,100	1, —	1, —	1, —	1, —	1,067	1,249	1,220	1,347
,083	1,046	1,099	1,092	1, —	1, —	1, —	1, —	1,307	1,297	1,475	1,553
,895	0,883	0,857	0,817	0,937	0,801	0,820	0,742	1, —	1, —	1, —	1, —
,828	0,807	0,745	0,703	0,765	0,771	0,678	0,644	1, —	1, —	1, —	1, —

M = operai metalmeccanici

C = operai edili

I, II, III, IV indicano rispettivamente:

I = operai specializzati
II = operai qualificati
III = manovali specializzati
IV = manovali comuni

PROSPETTO A

Graduatoria delle posizioni delle 7 città secondo le grandezze decrescenti dei salari, prezzi, tempi e poteri di acquisto come risulta dai rispettivi indici. La graduatoria secondo i prezzi servirà di base in quanto non varia con le categorie e classi.

	MI				MII			MIII			MIV			CI			CII			CIII			CIV		
	I_P	S	I_t	I_A	S	I_t	I_A	S	I_t	I_A	S	I_t	I_A	S	I_t	I_A	S	I_t	I_A	S	I_t	I_A	S	I_t	I_A
FORMULA 1ª																									
Messina . .	1	4	1	7	3	4	4	3	4	4	2	4	3	6	1	7	6	1	7	6	1	6	4,5	3	5
Milano . .	2	1	7	1	1	7	2	1	7	1	1	6	1	1	7	1	1	7	1	1	7	1	1	7	1
Salerno . .	3	5	4	4	6	2	6	5	3	5	5	2	5	5	3	5	4,5	3	5	4	4	4	3	5	3
Campobasso	4	2	5	3	5	3	5	4	5	3	4	3	4	7	2	6	7	2	6	7	1	7	6	2	6
Lecce . . .	5	6	2	6	7	1	7	6	2	6	6	1	6	4	4	4	3	5	3	5	3	5	7	1	7
Bolzano . .	6	3	6	2	2	6	1	2	6	2	3	5	2	2	6	2	2	6	2	3	5	3	2	6	2
Sassari . .	7	7	3	5	4	5	3	7	1	7	—	—	—	3	5	3	4,5	4	4	2	2	2	4,5	4	4
FORMULA 2ª																									
Messina . .	1	4	1	7	3	4	4	3	4	4	2	4	3	6	2	6	6	1	7	6	2	6	4,5	3	5
Milano . .	2	1	7	1	1	7	1	1	7	1	1	6	1	1	7	1	1	7	1	1	7	1	1	7	1
Bolzano . .	3	3	5	3	2	6	2	2	6	2	3	5	2	2	5	3	2	6	2	3	5	3	2	6	2
Salerno . .	4	5	4	4	6	2	6	5	3	5	5	2	5	5	3	5	4,5	3	5	4	4	4	3	5	3
Campobasso	5	2	6	2	5	3	5	4	5	3	4	3	4	7	1	7	7	2	6	7	1	7	6	1	7
Lecce . . .	6	6	3	5	7	1	7	6	2	6	6	1	6	4	4	4	3	5	3	5	3	5	7	2	6
Sassari . .	7	7	2	6	4	5	3	7	1	7	—	—	—	3	6	2	4,5	4	4	2	6	2	4,5	4	4
FORMULA 3ª																									
Messina . .	1	4	1	7	3	4	4	3	4	4	2	4	3	6	2	6	6	1	7	6	2	6	5	3	5
Milano . .	2	1	7	1	1	6	2	1	6	2	1	6	1	1	5	3	1	6	2	1	7	1	1	7	1
Salerno . .	3	5	3	5	6	2	6	5	3	5	5	2	5	5	3	5	4,5	3	5	4	4	4	3	4	4
Campobasso	4	2	5	3	5	3	5	4	5	3	4	3	4	7	1	7	7	2	6	7	1	7	6	1	7
Lecce . . .	5	6	2	6	7	1	7	6	2	6	6	1	6	4	4	4	3	4	4	5	3	5	7	2	6
Bolzano . .	6	3	6	2	2	7	1	2	7	1	3	5	2	2	7	1	2	7	1	3	5	3	2	6	2
Sassari . .	7	7	4	4	4	5	3	7	1	7	—	—	—	3	6	2	4,5	5	3	2	6	2	4	5	3
FORMULA 4ª																									
Messina . .	1	4	1	7	3	4	4	3	4	4	2	4	3	6	2	6	6	1	7	6	2	6	5	3	5
Milano . .	2	1	7	1	1	6	2	1	7	1	1	6	1	1	7	1	1	7	1	1	7	1	1	7	1
Salerno . .	3	5	4	4	6	2	6	5	3	5	5	2	5	5	3	5	4,5	3	5	4	4	4	3	4	4
Campobasso	4	2	6	2	5	3	5	4	5	3	4	3	4	7	1	7	7	2	6	7	1	7	6	2	6
Lecce . . .	5	6	2	6	7	1	7	6	2	6	6	1	6	4	4	4	3	5	3	5	3	5	7	1	7
Bolzano. .	6	3	5	3	2	7	1	2	6	2	3	5	2	2	6	2	2	6	2	3	5	3	2	6	2
Sassari . .	7	7	3	5	4	5	3	7	1	7	—	—	—	3	5	3	4,5	4	4	2	6	2	4	5	3

Prospetto B

Indici di cograduazione fra i poteri di acquisto e prezzi nelle 4 formule adoperate. Ad ogni graduatoria (per città) di poteri di acquisto di una classe di una certa categoria ottenuta con una formula si contrappone la graduatoria della stessa classe della stessa categoria ottenuta su ogni altra formula.

		Formula 1ª				Formula 2ª				Formula 3ª				Formula 4ª			
		I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
Formula 1ª . .	M	1, — *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*	0,833 *0,750*	0,917 *0,750*	1, — *0,750*	1, — *0,750*	0,917 *1, —*	1, — *1, —*	0,917 *1, —*	1, — *1, —*	0,917 *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*
	C	1, — *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*	0,833 *0,750*	1, — *0,750*	1, — *0,750*	0,917 *0,750*	0,750 *1, —*	0,833 *1, —*	1, — *1, —*	0,833 *1, —*	0,917 *1, —*	1, — *1, —*	1, — *1, —*	0,917 *1, —*
Formula 2ª . .	M	0,833 *0,750*	0,917 *0,750*	1, — *0,750*	1, — *0,750*	1, — *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*	0,750 *0,750*	0,958 *0,750*	0,917 *0,750*	1, — *0,750*	0,917 *0,750*	0,917 *0,750*	1, — *0,750*	1, — *0,750*
	C	0,833 *0,750*	1, — *0,750*	1, — *0,750*	0,917 *0,750*	1, — *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*	0,833 *0,750*	0,833 *0,750*	1, — *0,750*	0,917 *0,750*	0,917 *0,750*	1, — *0,750*	1, — *0,750*	0,833 *0,750*
Formula 3ª . .	M	0,917 *1, —*	1, — *1, —*	0,917 *1, —*	1, — *1, —*	0,750 *0,750*	0,958 *0,750*	0,917 *0,750*	1, — *0,750*	1, — *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*	0,833 *1, —*	1, — *1, —*	0,917 *1, —*	1, — *1, —*
	C	0,750 *1, —*	0,833 *1, —*	1, — *1, —*	0,833 *1, —*	0,833 *0,750*	0,833 *0,750*	1, — *0,750*	0,917 *0,750*	1, — *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*	0,833 *1, —*	0,833 *1, —*	1, — *1, —*	0,917 *1, —*
Formula 4ª . .	M	0,917 *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*	0,917 *0,750*	0,917 *0,750*	1, — *0,750*	1, — *0,750*	0,833 *1, —*	1, — *1, —*	0,917 *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*
	C	0,917 *1, —*	1, — *1, —*	1, — *1, —*	0,917 *1, —*	0,917 *0,750*	1, — *0,750*	1, — *0,750*	0,833 *0,750*	0,833 *1, —*	0,833 *1, —*	1, — *1, —*	0,917 *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*	1, — *1, —*

M = Metalmeccanici
C = Edili

I = operaio specializzato
II = operaio qualificato
III = manovale specializzato
IV = manovale comune

potere di acquisto = carattere normale
prezzi = » corsivo

PROSPETTO C

		MI			MII			MIII			MIV			CI			CII			CIII			CIV		
	I_p	S	I_t	I_A	S	I_t	I_A	S	I_t	I_A	S	I_t	I_A	S	I_t	I_A	S	I_t	I_A	S	I_t	I_A	S	I_t	I_A
Messina. .	1	4	1	7	3	4	4	3	4	4	2	4	3	6	2	6	6	1	7	6	2	6	4,5	3	5
Milano . .	2	1	7	1	1	6	2	1	7	1	1	6	1	1	7	1	1	7	1	1	7	1	1	7	1
Salerno . .	3	5	4	4	6	2	6	5	3	5	5	2	5	5	3	5	4	3	5	4	4	4	3	4,5	3,5
Campobasso	4	2	5,5	2,5	5	3	5	4	5	3	4	3	4	7	1	7	7	2	6	7	1	7	6	2	6
Lecce. . . .	5	6	2	6	7	1	7	6	2	6	6	1	6	4	4	4	3	5	3	5	3	5	7	1	7
Bolzano. .	6	3	5,5	2,5	2	7	1	2	6	2	3	5	2	2	5,5	2,5	2	6	2	3	5,5	2,5	2	6	2
Sassari . .	7	7	3	5	4	5	3	7	1	7	—	—	—	3	5,5	2,5	5	4	4	2	5,5	2,5	4,5	4,5	3,5

PROSPETTO D

Potere d'acquisto

	MI	MII	MIII	MIV	CI	CII	CIII	CIV		MI	MII	MIII	MIV	CI	CII	CIII	CIV
Milano	1	2	1	1	1	1	1	1	Campobasso. .	3	5	3	4	7	6	7	6
Bolzano. . . .	2	1	2	2	2,5	2	2,5	2	Messina. . . .	7	4	4	3	6	7	6	5
Sassari	5	3	7	—	2,5	4	2,5	3,5									
Salerno. . . .	4	6	5	5	5	5	4	3,5	Lecce.	6	7	6	6	4	3	5	7

PROSPETTO E

Tabella degli Indici di cograduazione tra le singole classi estratte dal prospetto D

	II				*I*			
	MI	MII	MIII	MIV	CI	CII	CIII	CIV
MI. . . .	1, —	0,333	0,625	0,556	0,375	0,500	0,500	0,583
MII . . .	0,333	1, —	0,458	0,889	0,542	0,375	0,542	0,625
MIII . . .	0,625	0,458	1, —	0,778	0,250	0,333	0,250	0,418
MIV . . .	0,556	0,889	0,778	1, —	0,444	0,444	0,444	0,667
CI	0,375	0,542	0,250	0,444	1, —	0,792	0,917	0,625
CII. . . .	0,500	0,375	0,333	0,444	0,792	1, —	0,625	0,542
CIII . . .	0,500	0,542	0,250	0,444	0,917	0,625	1, —	0,750
CIV . . .	0,583	0,625	0,418	0,667	0,625	0,542	0,750	1, —
	III				*IV*			

Montecatini

Società Generale per l'Industria Mineraria e Chimica

Alcune produzioni fondamentali:

Pirite - Zolfo - Lignite - Marmi grezzi e lavorati - Fertilizzanti azotati - Fertilizzanti fosfatici - Anticrittogamici e insetticidi - Acido solforico ed altri acidi per l'industria - Soda, cloro e derivati - Prodotti tartarici - Sali potassici - Solventi - Prodotti per la flottazione - Prodotti chimici vari per l'industria - Coke metallurgico - Catrame e derivati - Colle e gelatine - Pigmenti, vernici e smalti - Coloranti - Resine sintetiche - Farmaceutici - Esplosivi da mina e polveri da caccia - Olii lubrificanti e grassi - Fibre sintetiche - Filati e manufatti di juta - Alluminio, zinco e loro leghe - Piombo

MILANO - Via F. Turati, 18